Appunti di calcolo delle probabilità e statistica matematica

Francesca Biagini, Massimo Campanino [1]

6 giugno 2003

[1] Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato, 5, 40127 Bologna, Italia. Email: biagini@dm.unibo.it,campanin@dm.unibo.it

Indice

Capitolo 1

I numeri aleatori

1.1 Preliminari

Un *Numero Aleatorio* X è una quantità ben definita, ma non necessariamente nota, che si determina, ad esempio, effettuando un esperimento. Si possono conoscere i valori possibili, ovvero i valori che il numero aleatorio può assumere. Se X è il numero aleatorio, I(X) denota l'insieme dei valori possibili.

Esempio 1.1 Siano X, Y due numeri aleatori rappresentanti i risultati del lancio di una moneta e di un dado. Indicando croce con 0 e testa con 1, si ottiene

$$\mathrm{I(X)} = \{0,1\}$$
$$\mathrm{I(Y)} = \{1,2,3,4,5,6\}$$

Un numero aleatorio X si dice:

- *superiormente limitato* se l'insieme dei valori possibili I(X) è superiormente limitato ($\sup I(X) < +\infty$);
- *inferiormente limitato* se l'insieme dei valori possibili I(X) è inferiormente limitato ($\inf I(X) > -\infty$);
- *limitato* se l'insieme dei valori possibili I(X) è sia superiormente che inferiormente limitato ($\sup I(X) < +\infty$, $\inf I(X) > -\infty$).

Dati X e Y numeri aleatori, si definisce I(X,Y) l'insieme delle coppie possibili. In generale, si indica con I($X_1, ..., X_n$) l'insieme delle n-uple possibili.

X e Y si dicono *logicamente indipendenti* se

$$I(X, Y) = I(X) \times I(Y).$$

dove $I(X) \times I(Y)$ indica il prodotto cartesiano fra l'insieme dei valori possibili $I(X)$ e l'insieme dei valori possibili $I(Y)$.

Esempio 1.2 (di non indipendenza logica) Si considerino due estrazioni senza reimbussolamento di numeri da 1 a 90 (gioco del lotto), siano X_1 ed X_2 due numeri aleatori che rappresentano rispettivamente il risultato della prima e della seconda estrazione. L'insieme dei valori possibili è allora

$$I(X,Y) = \{(i,j)|1 \le i \le 90, i \neq j\}$$

Chiaramente, $I(X,Y) \neq I(X) \times I(Y)$ perché $I(X,Y)$ non contiene le coppie del tipo (i,i), $i \in \{1,...,90\}$. I due numeri aleatori X e Y non sono quindi logicamente indipendenti.

I numeri aleatori $X_1,...,X_n$ si dicono *logicamente indipendenti* se

$$I(X_1,...,X_n) = I(X_1) \times ... \times I(X_n).$$

Con i numeri aleatori si possono effettuare le usuali operazioni aritmetiche. Inoltre, si definiscono le seguenti operazioni:

1. $X \vee Y = \max(X,Y)$;
2. $X \wedge Y = \min(X,Y)$;
3. $\tilde{X} = 1 - X$

Tali operazioni hanno le seguenti proprietà:

1. Proprietà distributiva

(1) $$X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z)$$

(2) $$X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)$$

2. Proprietà associativa

(3) $$X \vee (Y \vee Z) = (X \vee Y) \vee Z$$

(4) $$X \wedge (Y \wedge Z) = (X \wedge Y) \wedge Z$$

3. Proprietà commutativa

(5) $$X \vee Y = Y \vee X$$

(6) $$X \wedge Y = Y \wedge X$$

4. Proprietà connesse alla ˜

(7) $$\tilde{\tilde{X}} = X$$

(8) $$(X \vee Y)\tilde{} = \tilde{X} \wedge \tilde{Y}$$

(9) $$(X \wedge Y)\tilde{} = \tilde{X} \vee \tilde{Y}$$

1.2 Eventi

Un caso particolare di numero aleatorio è dato dagli eventi. Un evento E è un numero aleatorio tale che $I(E) \subset \{0,1\}$. Nel caso di eventi, dati due eventi E e F, $E \vee F$ si dice *somma logica* e $E \wedge F$ *prodotto logico*. Si verifica facilmente che:

1. $E \vee F = E + F - EF$;

2. $E \wedge F = EF$.

Dato un evento E, si definisce *complementare* di E l'evento

$$\tilde{E} = 1 - E$$

Si ha che $\tilde{\tilde{E}} = E$.

Dalle proprietà della ˜ abbiamo

$$(E \vee F)\tilde{} = \tilde{E} \wedge \tilde{F} = (1-E)(1-F) = 1 - E - F + EF,$$

da cui segue

$$E \vee F = E + F - EF.$$

Analogamente

$$(E \vee F \vee G)\tilde{} = \tilde{E} \wedge \tilde{F} \wedge \tilde{G} = (1-E)(1-F)(1-G) = 1 - E - F - G + EF + EG + FG - EFG,$$

da cui segue

$$E \vee F \vee G = E + F + G - EF - EG - FG + EFG.$$

Altre due operazioni fra eventi sono:

differenza: $E \setminus F = E - EF$;

differenza simmetrica: $E \triangle F = (E \setminus F) \vee (F \setminus E)$.

D'ora in avanti useremo il simbolo $\vdash$ per dire che la proposizione che segue è sicuramente vera. Per esempio, $\vdash X \leq Y$ se $\sup I(X) \leq \inf I(Y)$.

Esiste una corrispondenza con le operazioni insiemistiche. Si pone

$$E \subset F \text{ se} \vdash E \leq F,$$

$$E \equiv F \text{ se } E \subset F \text{ e } F \subset E.$$

Definizione 1.3 *Si definiscono le seguenti proprietà:*

1. ***Incompatibilità**: E, F si dicono incompatibili se $\vdash EF = 0$;*

2. ***Esaustività**: E_1,...,E_n si dicono esaustivi se $\vdash E_1 + ... + E_n \geq 1$;*

3. ***Partizione**: E_1,...,E_n si dicono una partizione se $\vdash E_1 + ... + E_n = 1$ (esaustivi e incompatibili).*

Esempio 1.4 Un evento E ed il suo complementare $\tilde{E}$ sono una partizione.

Siano $E_1, \ldots, E_n$ eventi, per trovare una partizione si usa il metodo dei costituenti. Si definisce *costituente* di $E_1, \ldots, E_n$ l'evento

$$Q = E_1^* \cdots E_n^*$$

dove E_i^* è definito nel seguente modo

$$E_i^* = \begin{cases} E_i \\ \tilde{E}_i \end{cases}$$

In generale, non tutti i costituenti sono possibili. Sono possibili tutti i costituenti solo quando gli E_i sono logicamente indipendenti. I costituenti possibili sono una partizione. Infatti

$$1 = (E_1 + \tilde{E}_1) \ldots (E_n + \tilde{E}_n) = \sum_{Q \text{costituente}} Q$$

Dalla somma si possono escludere tutti i costituenti impossibili.

Se $E_1, ..., E_n$ sono una partizione, allora i costituenti possibili sono:

$$\begin{gathered} E_1 \tilde{E}_2 \cdots \tilde{E}_n, \\ \tilde{E}_1 E_2 \tilde{E}_3 \cdots \tilde{E}_n, \\ \cdots, \\ \tilde{E}_1 \cdots \tilde{E}_{n-1} E_n. \end{gathered}$$

In questo caso, il costituente $\tilde{E}_1 \cdots \tilde{E}_n$ non è possibile.

Definiamo ora quando un evento è logicamente indipendente da altri eventi. I costituenti sono classificabili nel seguente modo rispetto ad un dato evento E:

I tipo $Q \subset E$;

II tipo $Q \subset \tilde{E}$;

III tipo altrimenti.

E è *logicamente dipendente* da E_1,...,E_n se tutti i costituenti di E_1,...,E_n sono del primo o del secondo tipo.

E è *logicamente indipendente* da E_1,...,E_n se tutti i costituenti di E_1,...,E_n sono del terzo tipo.

Altrimenti E si dice *semidipendente*.

Se E è logicamente dipendente da E_1,...,E_n, si può scrivere

$$E = \sum_{\substack{Q \text{ di I tipo} \\ Q \subset E}} Q$$

Esempio 1.5 Consideriamo due eventi E_1, E_2. L'evento somma logica $E_1 \vee E_2$ si può scrivere come

$$E_1 \vee E_2 = E_1E_2 + \tilde{E}_1E_2 + E_1\tilde{E}_2$$

In generale se un evento E è logicamente dipendente da E_1,...,E_n se e solo se E si può scrivere come $E = \Phi(E_1, ..., E_n)$ per qualche funzione Φ.

Esempio 1.6 Supponiamo di effettuare cinque lanci di una moneta. Sia E_i l'evento che corrisponde all'esito "testa" all'i-esimo lancio. Posto $Y = E_1 + E_2 + E_3 + E_4 + E_5$, considero l'evento

$$E = (Y \geq 3).$$

E è semidipendente dai primi tre eventi. Infatti

I tipo: $E_1E_2E_3 \subset E$;

II tipo: $\tilde{E}_1\tilde{E}_2\tilde{E}_3 \subset \tilde{E}$;

III tipo: $\tilde{E}_1E_2\tilde{E}_3$.

1.3 La previsione

Dato un numero aleatorio X, cerchiamo un valore certo che esprima la nostra valutazione su X. In termini economici se pensiamo a X come a un guadagno aleatorio, vogliamo scegliere un guadagno certo che riteniamo equivalente a X.

Seguendo l'impostazione di de Finetti (vedi ad esempio: B. de Finetti "Teoria delle Probabilità", Einaudi), definiamo in modo operativo la previsione $\mathbb{P}(X)$[1] che un individuo assegna ad un numero aleatorio X.

Esistono due modi operativi equivalenti per definire la previsione:

1. **Metodo della scommessa**: si pensa X come il guadagno (o le perdite) derivante da una scommessa. La previsione $\mathbb{P}(X)$ è allora il guadagno certo che si giudica equivalente alla quantità aleatoria X.

   Posto $\mathbb{P}(X) = \mathbb{E}[X] = \bar{x}$, si accetta una scommessa pari a

   $$\lambda(X - \bar{x})$$

   dove $\lambda \in \mathbb{R}$ è un coefficiente di proporzionalità. Il corrispondente criterio di coerenza è che non si possa scegliere $\bar{x}$ in modo che ci sia una perdita certa. Nella finanza matematica questo prende il nome di *Principio di Non Arbitraggio*.

2. **Metodo della penalità**: si suppone di dover pagare una penalità pari a

   $$-\lambda(X - \bar{\bar{x}})^2$$

[1] Prende anche il nome di *media*, *attesa* o *speranza*, e si indica anche con $\mathbb{E}(X)$.

dove $\lambda \in \mathbb{R}$ è un coefficiente di proporzionalità. Si cerca $\bar{\bar{x}}$ che minimizzi le perdite (criterio di *minima distanza*). Tale $\bar{\bar{x}}$ si dice previsione $\mathbb{P}(X)$ del numero aleatorio X. Anche qui vi è un criterio di coerenza: non deve esistere un valore $\bar{\bar{x}}'$ tale che la corrispondente penalità sia sicuramente minore.

Proposizione 1.7 (Proprietà della previsione) *La previsione ha le seguenti proprietà:*

1. *Monotonia:* $\inf I(X) \leq \mathbb{P}(X) \leq \sup I(X)$*;*
2. *Linearità: se* $X = \alpha_1 X_1 + \cdots + \alpha_n X_n$*, allora* $\mathbb{P}(X) = \alpha_1 \mathbb{P}(X_1) + \cdots + \alpha_n \mathbb{P}(X_n)$.

DIMOSTRAZIONE.

1. *Monotonia*: Supponiamo $\lambda > 0$. $\bar{x}$ deve essere tale che non si abbia un guadagno certo od una perdita certa.

$$\bar{x} < \inf I(X) \longrightarrow \text{si vince certamente}$$
$$\bar{x} > \sup I(X) \longrightarrow \text{si perde certamente}$$

   Se $\lambda < 0$, vale il viceversa. Ne segue che

$$\inf I(X) \leq \bar{x} \leq \sup I(X).$$

   Tale proprietà si dimostra allo stesso modo se si usa il secondo criterio.

2. *Linearità*: Per la dimostrazione, si procede utilizzando il principio di Non Arbitraggio.

   Consideriamo il numero aleatorio $Z = X + Y$.

   Posto $\bar{z} = \mathbb{P}(Z)$, $\bar{x} = \mathbb{P}(X)$, $\bar{y} = \mathbb{P}(Y)$, sia G il guadagno

$$G = c_1(X - \bar{x}) + c_2(Y - \bar{y}) + c_3(Z - \bar{z}) =$$

$$= (c_1 + c_3)X + (c_2 + c_3)Y - c_1\bar{x} - c_2\bar{y} - c_3\bar{z}$$

   Scegliendo c_1, c_2, c_3 in modo tale da annullare la parte aleatoria

$$c_1 = c_2 = -c_3,$$

   si ottiene il guadagno complessivo: $G = c_3(\bar{x} + \bar{y} - \bar{z})$.

   Per evitare vincite o perdite certe, dovrà essere $\bar{x} + \bar{y} - \bar{z} = 0$, ovvero $\bar{z} = \bar{x} + \bar{y}$.

   Se si procede invece con il secondo criterio, si è sottoposti ad una penalità (guadagno negativo)

$$-[(X - \bar{x})^2 + (Y - \bar{y})^2 + (Z - \bar{z})^2] = -[(X - \bar{x})^2 + (Y - \bar{y})^2 + (X + Y - \bar{z})^2]$$

   Si cerca il punto P di coordinate $(\bar{x}, \bar{y}, \bar{z})$ che ha minima distanza dal piano $z = x + y$. Per ogni punto P, la proiezione ortogonale P' di P sul piano $z = x + y$ ha distanza più

piccola di P dal piano. In base al principio di coerenza dovrà essere $P = P'$, ovvero che P deve appartenere al piano $z = x + y$. Ne segue che $\bar{z} = \bar{x} + \bar{y}$.

Analogamente, per $Z = \alpha X$, $\alpha \in \mathbb{R}$, si ottiene $\bar{z} = \alpha \bar{x}$.

In generale, se $X = \alpha_1 X_1 + \cdots + \alpha_n X_n$, allora

$$\mathbb{P}(X) = \alpha_1 \mathbb{P}(X_1) + \cdots + \alpha_n \mathbb{P}(X_n)$$

□

La proprietà di monotonia si può descrivere anche nel seguente modo:

$\vdash X \geq c \Longrightarrow \mathbb{P}(X) \geq c$;

Se $c_1 \leq c_2$, $\vdash c_1 \leq X \leq c_2 \Longrightarrow c_1 \leq \mathbb{P}(X) \leq c_2$

$\vdash X = c \Longrightarrow \mathbb{P}(X) = c$.

1.4 Probabilità di eventi

Nel caso di un evento, la previsione $\mathbb{P}(E)$ si chiama *probabilità* di E. Dalle proprietà di monotonia e linearità, segue che:

1. la probabilità di un evento è un numero compreso fra 0 ed 1, ovvero $0 \leq \mathbb{P}(E) \leq 1$.
2. $E \equiv 0 \Longrightarrow \mathbb{P}(E) = 0$
3. $E \equiv 1 \Longrightarrow \mathbb{P}(E) = 1$

Quando $E \equiv 1$, E si dice l'*evento certo*.
Si ha che:

somma logica: $\mathbb{P}(E_1 \vee E_2) = \mathbb{P}(E_1 + E_2 - E_1 E_2) \leq \mathbb{P}(E_1 + E_2)$;

somma: $\mathbb{P}(E_1 + E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$.

Le due previsioni coincidono *se e solo se* E_1 e E_2 sono incompatibili. In generale

$$\mathbb{P}(E_1 \vee E_2) \leq \mathbb{P}(E_1 + E_2)$$

Per la monotonia della previsione si ha infatti che

$$\vdash E_1 + E_2 - E_1 \vee E_2 \geq 0 \Longrightarrow \mathbb{P}(E_1 + E_2 - E_1 \vee E_2) \geq 0$$

Per una partizione

$$\vdash E_1 + \cdots + E_n = 1 \Longrightarrow \sum \mathbb{P}(E_i) = 1$$

La funzione che assegna agli eventi di una partizione le loro probabilità si dice *distribuzione di probabilità*. Se E dipende logicamente da una partizione di eventi $\{E_1, \ldots, E_n\}$ possiamo trovare la probabilità di E a partire da quella degli E_i.

$$\mathbb{P}(E) = \sum_{E_i \subset E} \mathbb{P}(E_i)$$

Vediamo ora un metodo operativo per calcolare la previsione. Sia X un numero aleatorio con $I(X) = \{x_1, ..., x_n\}$ e sia $E_i := (X = x_i)$. Si ha che:

1.
$$\mathbb{P}(X) = \mathbb{P}(X(E_1 + \cdots + E_n)) =$$
$$= \mathbb{P}(XE_1) + \cdots + \mathbb{P}(XE_n) =$$
$$= \textstyle\sum_{i=1}^n \mathbb{P}(XE_i) = \sum_{i=1}^n \mathbb{P}(x_i E_i) = \sum_{i=1}^n x_i \mathbb{P}(E_i) =$$
$$= \textstyle\sum_{i=1}^n x_i \mathbb{P}(X = x_i)$$

Basta infatti notare che XE_i è un numero aleatorio che assume il valore x_i oppure 0. L'uguaglianza $\mathbb{P}(x_i E_i) = x_i \mathbb{P}(E_i)$ è una conseguenza della proprietà di linearità della previsione.

2. In generale, se $I(X)$ è finito e $\Phi : \mathbb{R} \to \mathbb{R}$ vale che
$$\mathbb{P}(\Phi(X)) = \sum_{i=1}^n \Phi(x_i)\mathbb{P}(X = x_i)$$

La dimostrazione è analoga a quella del punto 1.

Esempio 1.8 Sia X il numero rappresentante il risultato del lancio di un dado. Se ogni faccia ha la stessa probabilità di uscire, la previsione di X è data da:

$$\mathbb{P}(X) = \frac{\sum i}{6} = \frac{6 \cdot 7}{6 \cdot 2} = \frac{7}{2}$$

Esempio 1.9 Sia X il numero aleatorio che rappresenta il il risultato del lancio di una moneta simmetrica. Indicando con $I(X) = \{0, 1\}$, si ottiene che la previsione di X è data da:

$$\mathbb{P}(X) = \frac{1}{2}$$

1.5 Partizioni in eventi equiprobabili

In alcune situazioni, per ragioni di simmetria, è naturale attribuire la stessa probabilità a tutti gli eventi di una partizione, come nel caso dei giochi d'azzardo. Se $E_1, \ldots, E_n$ sono gli eventi di una partizione con *distribuzione uniforme*, vale che

$$\mathbb{P}(E_i) = \frac{1}{n}$$

Sia E un evento che dipende logicamente dalla partizione $E_1, \ldots, E_n$. La previsione di E è data da:

$$\mathbb{P}(E) = \mathbb{P}\left(\sum_{E_i \subset E} E_i\right) = \frac{\sharp\{i | E_i \subset E\}}{n}$$

si ottiene dunque la nota formula

$$\mathbb{P}(E) = \frac{\sharp\, casi\ favorevoli}{\sharp\, casi\ possibili}$$

Tale identità è valida unicamente nel caso in cui gli eventi della partizione sono valutati equiprobabili.

Esempio 1.10 Si effettuano n lanci di una moneta equilibrata. Sia X il numero aleatorio che rappresenta il numero di teste che si ottengono considerando n lanci. Sia E_i l'evento corrispondente all'uscita di una *testa* all'i-esimo lancio. Considero l'evento

$$E := (X = k) = \sum_{Q \subset E} Q$$

dove $Q = E_1^* \ldots E_n^*$ sono i costituenti degli eventi $E_1 \ldots E_n$; tali costituenti determinano una partizione e sono tutti possibili in quanto gli E_i sono tutti logicamente indipendenti.

$$\sharp\, casi\ possibili = 2 \cdot \ldots \cdot 2 = 2^n \Rightarrow \mathbb{P}(Q) = \frac{1}{2^n}$$

$$\sharp\, casi\ favorevoli = \binom{n}{k}$$

Ne segue che

$$\mathbb{P}(E) = \binom{n}{k} \frac{1}{2^n}$$

Si cerca per quale k la probabilità $\mathbb{P}(X = k)$ è massima. Si ottiene che:

n pari: Il massimo valore di $\mathbb{P}(X = k)$ si ha per $k = \frac{n}{2}$;

n dispari: Il massimo valore di $\mathbb{P}(X = k)$ si ha per $k = \frac{n-1}{2}$ e per $k = \frac{n+1}{2}$.

Esempio 1.11 Si fanno n estrazioni con reimbussolamento da un'urna con H palline bianche e $N-H$ palline nere. Sia X il numero aleatorio che conta il numero di palline bianche estratte. Si calcola

$$\mathbb{P}(X = k) = \frac{\sharp\, casi\ favorevoli}{\sharp\, casi\ possibili}$$

dove il numero di *casi possibili* è pari a N^n ed il numero di *casi favorevoli* è pari a

$$\binom{n}{k} H^k (N - H)^{n-k}.$$

Si può pensare un costituente come una sequenza di palline bianche e nere. I casi favorevoli sono quelli in cui tale sequenza presenta una pallina bianca in k posizioni; per ciascuna di queste posizioni si può scegliere tra H palline bianche, essendo le estrazioni con reimbussolamento.

Considerando invece delle estrazioni senza reimbussolamento, il numeri di casi possibili è dato da

$$\binom{N}{n}$$

Possiamo infatti non tener conto dell'ordine in quanto l'evento considerato non dipende dall'ordine di estrazione delle n palline. Il numero di casi favorevoli è dato da

$$\binom{H}{k}\binom{N-H}{n-k}$$

Si devono infatti scegliere k palline fra le H bianche senza tener conto dell'ordine ed $n-k$ palline fra le $N-H$ nere senza tener conto dell'ordine. Alternativamente avremmo potuto tener conto dell'ordine (ovviamente sia nel conteggio dei casi favorevoli sia in quello dei casi possibili), ottenendo lo stesso risultato.

1.6 Probabilità e previsione subordinata

Si tratta della probabilità (e della previsione) subordinata (o condizionata) al verificarsi di un dato evento. Sia X un numero aleatorio ed H un evento. Per definire la previsione subordinata, si utilizzano due metodi operativi.

1. *Metodo della scommessa*:

   La scommessa vale quando H si verifica, altrimenti è annullata e quindi il guadagno è uguale a 0. Si sceglie $\bar{x}$ sapendo che si può essere sottoposti ad una scommessa con un guadagno:

   $$G = \lambda H(x - \bar{x})$$

   dove $\lambda \in \mathbb{R}$ rappresenta un coefficiente di proporzionalità. $\bar{x}$ si dice *previsione subordinata* di X rispetto ad H e si denota con $\mathbb{P}(X|H)$.

2. *Metodo della penalità*:

   Anche qui la penalità viene inflitta se H si verifica. Si sceglie $\bar{\bar{x}}$ sapendo di dover pagare una penalità

   $$p = -H(X - \bar{\bar{x}})^2$$

   $\bar{\bar{x}}$ è *previsione subordinata* di X rispetto ad H e si denota con $\mathbb{P}(X|H)$.

Considero l'insieme dei valori possibili $I(X|H)$ di X dato H. La previsione subordinata ha le stesse proprietà della previsione, ovvero:

- $I(X|H) \subset I(X)$
- $\inf I(X|H) \leq \mathbb{P}(X|H) \leq \sup I(X|H)$
- $\mathbb{P}(X+Y|H) = \mathbb{P}(X|H) + \mathbb{P}(Y|H)$
- $\mathbb{P}(\lambda X|H) = \lambda \mathbb{P}(X|H)$

1.7 Formula delle probabilità composte

Vale la *formula delle probabilità composte*

$$\mathbb{P}(XH) = \mathbb{P}(H)\mathbb{P}(X|H)$$

Per dimostrarla, si pongano $z = P(XH)$, $x = P(H)$ e $y = P(X|H)$. Utilizzando il metodo della scommessa, si ottiene:

$$G = c_1(H - x) + c_2H(X - y) + c_3(XH - z)$$

$$= H(c_1 + (c_2 + c_3)X - c_2y) - c_1x - c_3z$$

Ponendo $c_2 = -c_3$ e $c_1 = c_2y$ si ottiene

$$G \ = \ -c_1x - c_3z \ = \ c_2(z - xy)$$

Avendo annullato la parte aleatoria, per non avere arbitraggi, ovvero guadagni o perdite certe, si dovrà avere:

$$z = xy$$

In modo analogo, è possibile dimostrare la formula con il metodo delle penalità.

Se $\mathbb{P}(H) > 0$, vale

$$\mathbb{P}(X|H) = \frac{\mathbb{P}(XH)}{\mathbb{P}(H)}$$

Se X è un evento, $X = E$, allora

$$\mathbb{P}(E|H) = \frac{\mathbb{P}(EH)}{\mathbb{P}(H)}$$

Casi particolari:

- $$E \subset H \ \ \Rightarrow \ \ \mathbb{P}(E|H) = \frac{\mathbb{P}(E)}{\mathbb{P}(H)}$$

- $$H \subset E, \text{ ovvero} I(E|H) = \{1\} \ \ \Rightarrow \ \ \mathbb{P}(E|H) = 1$$

- $$H \subset \tilde{E}, \text{ ovvero } I(E|H) = \{0\} \ \ \Rightarrow \ \ \mathbb{P}(E|H) = 0$$

1.7.1 Formula delle probabilità totali

Sia $H_1, \ldots, H_n$ una partizione e X un numero aleatorio. Vale che:

$$\mathbb{P}(X) = \sum_{i=1}^{n} \mathbb{P}(X|H_i)\mathbb{P}(H_i)$$

Infatti,

$$\mathbb{P}(X) = \mathbb{P}(X \cdot 1) = \mathbb{P}(X(H_1 + \ldots + H_n)) =$$
$$\mathbb{P}(XH_1 + XH_2 + \cdots + XH_n) = \sum_{i=1}^{n} \mathbb{P}(XH_i) = \sum_{i=1}^{n} \mathbb{P}(X|H_i)\mathbb{P}(H_i)$$

1.7.2 Formula di Bayes

Siano E, H due eventi tali che $\mathbb{P}(H) > 0$. Vale la *Formula di Bayes*

$$\mathbb{P}(E|H) = \frac{\mathbb{P}(H|E)\mathbb{P}(E)}{\mathbb{P}(H)}$$

Dalla formula della probabilità condizionata si ha che $\mathbb{P}(EH) = \mathbb{P}(H|E)\mathbb{P}(E)$. Quindi:

$$\mathbb{P}(E|H) = \frac{\mathbb{P}(EH)}{\mathbb{P}(H)} = \frac{\mathbb{P}(H|E)\mathbb{P}(E)}{\mathbb{P}(H)}$$

Esempio 1.12 Consideriamo un'urna di composizione ignota contenente N palline bianche e nere. Sia Y il numero aleatorio di palline bianche nell'urna.

Gli eventi $H_i = (Y = i)$ determinano una partizione. Sia E l'evento corrispondente all'estrazione di una pallina bianca. Si calcoli la probabilità di E e la probabilità che nell'urna vi siano i palline bianche se si è estratta una pallina bianca, ovvero se si è verificato l'evento E.

Si usa la formula delle probabilità totali per calcolare la probabilità di E

$$\mathbb{P}(E) = \sum_{i=0}^{N} \mathbb{P}(E|H_i)\mathbb{P}(H_i) = \sum_{i=0}^{N} \frac{i}{N}\mathbb{P}(H_i)$$

Se è nota la composizione dell'urna, la probabilità di E è data dal numero di palline bianche, cioè i casi favorevoli, diviso il numero totale delle palline, i casi possibili.

Supponiamo ora che non si conosca la composizione dell'urna. Se si assume che gli H_i siano equiprobabili, cioè che sia $\mathbb{P}(H_i) = \frac{1}{N+1}$, si ottiene:

$$\mathbb{P}(E) = \sum_{i=0}^{N} \frac{i}{N(N+1)} = \frac{1}{2}$$

Dalla formula di Bayes segue che

$$\mathbb{P}(H_i|E) = \frac{\mathbb{P}(E|H_i)\mathbb{P}(H_i)}{\mathbb{P}(E)} = \frac{\frac{i}{N}\frac{1}{N+1}}{\frac{1}{2}} = \frac{2i}{N(N+1)}$$

1.8 Correlazione tra eventi

Un evento E si dice *correlato positivamente* con H se

$$\mathbb{P}(E|H) > \mathbb{P}(E)$$

Analogamente, un evento E si dice *correlato negativamente* con H se

$$\mathbb{P}(E|H) < \mathbb{P}(E)$$

Se $\mathbb{P}(E|H) = \mathbb{P}(E)$, si dice che E non è correlato con H; in tal caso si dice anche che E ed H sono *stocasticamente indipendenti.*

In questo caso, l'informazione che H si è verificato non cambia la valutazione delle probabilità di E e viceversa. Se invece E è correlato positivamente con H, l'informazione che H si è verificato aumenta la valutazione della probabilità di E.

Se $\mathbb{P}(H) > 0$ e $\mathbb{P}(E) > 0$, si può dare una definizione simmetrica della correlazione. E e H si dicono

- *correlati positivamente* se $\mathbb{P}(EH) > \mathbb{P}(E)\mathbb{P}(H)$
- *correlati negativamente* se $\mathbb{P}(EH) < \mathbb{P}(E)\mathbb{P}(H)$
- *non correlati* se $\mathbb{P}(EH) = \mathbb{P}(E)\mathbb{P}(H)$

Se $\mathbb{P}(E)$ e $\mathbb{P}(H)$ sono entrambi positivi, la correlazione si dice simmetrica. Se E è correlata positivamente con H, si ha che $\tilde{E}$ è correlato negativamente con H

$$\mathbb{P}(\tilde{E}|H) = 1 - \mathbb{P}(E|H) < 1 - \mathbb{P}(E) = \mathbb{P}(\tilde{E})$$

Se E non è correlato con H, nemmeno $\tilde{E}$ lo è.

Esempio 1.13 Consideriamo un'urna con H palline bianche e $N - H$ palline nere; si effettuano due estrazioni. Si denotano con E_1, E_2 gli eventi in cui si estratta una pallina bianca rispettivamente alla prima ed alla seconda estrazione. Nel caso di estrazioni con reimbussolamento, si ottiene

$$\mathbb{P}(E_1) = \frac{H}{N}$$
$$\mathbb{P}(E_2) = \frac{H}{N}$$

Infatti la composizione dell'urna è la stessa sia alla prima che alla seconda estrazione. Si verifica subito che le due estrazioni sono indipendenti (come ci si aspettava!) in quanto

$$\mathbb{P}(E_1E_2) = \frac{H^2}{N^2} = \mathbb{P}(E_1)\mathbb{P}(E_2)$$

Se invece si effettuano le estrazioni senza reimbussolamento

$$\mathbb{P}(E_1) = \frac{H}{N}$$

$$\mathbb{P}(E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) + \mathbb{P}(E_2|\tilde{E}_1)\mathbb{P}(\tilde{E}_1) = \\ = \frac{H-1}{N-1}\frac{H}{N} + \frac{H}{N-1}(1-\frac{H}{N}) = \frac{H}{N}$$

Le probabilità delle due estrazioni sono dunque le stesse, ma i due eventi risultano correlati negativamente in quanto

$$\mathbb{P}(E_2|E_1) = \frac{H-1}{N-1} < \frac{H}{N} = \mathbb{P}(E_2)$$

per ogni $H < N$.

È possibile estendere la definizione di indipendenza anche al caso di un numero n, generico, di eventi. $E_1, \ldots, E_n$ si dicono *stocasticamente indipendenti* se per ogni scelta finita di indici $\{i_1, \ldots, i_k\}$ in $\{1, \ldots, n\}$ si ha che

$$(1) \quad \mathbb{P}(E_{i_1} \cdots E_{i_k}) = \mathbb{P}(E_{i_1}) \cdots \mathbb{P}(E_{i_k})$$

Non basta verificare la (1) solamente per le coppie!

Se $E_1, \ldots, E_n$ sono *stocasticamente indipendenti*, anche $E_1^*, \ldots, E_n^*$ sono *stocasticamente indipendenti* per ogni scelta possibile di E_i^* fra E_i ed $\tilde{E}_i$.

Sia $\mathcal{H} = \{H_1, \ldots, H_n\}$ una partizione; due eventi E_1, E_2 si dicono *stocasticamente indipendenti subordinatamente alla partizione* $\mathcal{H}$ se

$$\forall i = 1, \ldots n \;\; \mathbb{P}(E_1E_2|H_i) = \mathbb{P}(E_1|H_i)\mathbb{P}(E_2|H_i)$$

Esempio 1.14 Consideriamo un'urna con composizione incognita contenente N palline bianche e nere. Sia Y il numero aleatorio che rappresenta il numero di palline nell'urna. Si effettuano due estrazioni con reimbussolamento. Sia E_1 l'evento "esce una pallina bianca alla prima estrazione" e sia E_2 l'evento "esce una pallina bianca alla seconda estrazione".

Consideriamo la partizione $\mathcal{H}$ determinata dagli eventi

$$H_i = (Y = i) \quad i = 0, \ldots N$$

Si assume che $\mathbb{P}(H_i) = \frac{1}{N+1}$. Gli eventi E_1 ed E_2 sono stocasticamente indipendenti data $\mathcal{H}$, ovvero

$$\mathbb{P}(E_1E_2|H_i) = \mathbb{P}(E_1|H_i)\mathbb{P}(E_2|H_i)$$

per ogni $i = 0, \ldots N$.Infatti, se la composizione dell'urna è nota, le estrazioni con reimbussolamento non si influenzano reciprocamente. Ci si chiede se essi siano anche stocasticamente indipendenti. Si calcola

1. La probabilità della prima estrazione

$$\begin{aligned}\mathbb{P}(E_1) &= \sum_{i=0}^{N} \mathbb{P}(E_1|H_i)\mathbb{P}(H_i) \\ &= \frac{1}{N+1}\sum_{i=0}^{N}\frac{i}{N} \\ &= \frac{1}{N+1}\frac{N(N+1)}{2} \\ &= \frac{1}{2}\end{aligned}$$

2. La probabilità della seconda estrazione è pari a quella della prima perché si è già notato che $\mathbb{P}(E_1|H_i) = \mathbb{P}(E_2|H_i)$. Quindi $\mathbb{P}(E_2) = \mathbb{P}(E_1) = \frac{1}{2}$

3.

$$\begin{aligned}\mathbb{P}(E_1E_2) &= \sum_{i=0}^{N} \mathbb{P}(E_1E_2|H_i)\mathbb{P}(H_i) \\ &= \frac{1}{N+1}\sum_{i=0}^{N}\mathbb{P}(E_1|H_i)\mathbb{P}(E_2|H_i) \\ &= \frac{1}{N+1}\sum_{i=0}^{N}\frac{i^2}{N^2}\end{aligned}$$

Per calcolare $\sum_{i=0}^{N} k^2$ si utilizza il fatto che

$$(n+1)^3 - n^3 = 3n^2 + 3n + 1$$

Ne segue che $\sum_{i=0}^{N} k^2 = \frac{(N+1)^3}{3} - \frac{N(N+1)}{2} - \frac{(N+1)}{3}$ ed inoltre

$$\mathbb{P}(E_1E_2) = \frac{2N+1}{6N}$$

Tale probabilità tende ad $\frac{1}{3}$ per N che tende all'infinito. Quindi $\mathbb{P}(E_1E_2) \neq \mathbb{P}(E_1)\mathbb{P}(E_2)$ ovvero l'indipendenza stocastica rispetto ad una partizione non implica l'indipendenza stocastica.

1.8.1 L'indipendenza stocastica attraverso i costituenti

$E_1, \ldots, E_n$ sono stocasticamente indipendenti se per ogni costituente $Q = E_1^* \cdots E_n^*$ di $E_1, \ldots, E_n$, dove

$$E_i^* = \begin{cases} E_i \\ \tilde{E}_i \end{cases}$$

vale che

$$\mathbb{P}(Q) = \mathbb{P}(E_1^*) \cdots \mathbb{P}(E_n^*) \tag{10}$$

DIMOSTRAZIONE.

$\Rightarrow$) Sia $Q = E_1^* \cdots E_n^*$ un costituente di $E_1, \ldots, E_n$. Se si sviluppano i prodotti fra gli eventi, si ottiene che Q è dato da un polinomio Φ in n variabili di grado 1 in ogni variabile calcolato in $E_1, \ldots, E_n$, ovvero

$$E_1^* \cdots E_n^* = \Phi(E_1, \ldots, E_n)$$

Per esempio, se si considerano tre eventi E_1, E_2, E_3, il costituente $Q = \tilde{E}_1 E_2 E_3 = (1 - E_1)E_2E_3 = E_2E_3 - E_1E_2E_3$, ovvero si ottiene dal polinomio in tre variabili $\phi(x_1, x_2, x_3) = x_2x_3 - x_1x_2x_3$ calcolato in E_1, E_2, E_3.

Se gli E_i sono stocasticamente indipendenti, le probabilità dei prodotti si fattorizzano e si ottiene

$$\begin{aligned} \mathbb{P}(Q) &= \mathbb{P}\left(\Phi(E_1, \ldots, E_n)\right) \\ &= \Phi\left(\mathbb{P}(E_1), \ldots, \mathbb{P}(E_n)\right) \\ &= \mathbb{P}(E_1^*) \cdots \mathbb{P}(E_n^*) \end{aligned}$$

Ritornando all'esempio, si calcoli la probabilità di $Q = \tilde{E}_1 E_2 E_3$.

$$\begin{aligned} \mathbb{P}(Q) &= \mathbb{P}\left(\tilde{E}_1 E_2 E_3\right) \\ &= \mathbb{P}\left(E_2E_3 - E_1E_2E_3\right) \\ &= \mathbb{P}(E_2)\mathbb{P}(E_3) - \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3) = \Phi(\mathbb{P}(E_1), \mathbb{P}(E_2), \mathbb{P}(E_3)) \end{aligned}$$

$\Leftarrow$) Viceversa, si supponga che valga (10). Dati $E_1, \ldots, E_n$ essi sono stocasticamente indipendenti se e solo se per ogni scelta di indici $i_1, \ldots, i_k$ in $\{1, \ldots, n\}$ vale che

$$\mathbb{P}(E_{i_1} \cdots E_{i_k}) = \mathbb{P}(E_{i_1}) \cdots \mathbb{P}(E_{i_k})$$

Usando i costituenti, si ha che:

$$\begin{aligned} \mathbb{P}(E_{i_1} \cdots E_{i_k}) &= \mathbb{P}\left(\sum_{Q \subset E_{i_1} \cdots E_{i_k}} Q\right) \\ &= \mathbb{P}(E_{i_1}) \cdots \mathbb{P}(E_{i_k}) \cdot \sum \left(\mathbb{P}(E_{i_{k+1}}^*) \cdots \mathbb{P}(E_{i_N}^*)\right) \end{aligned}$$

La sommatoria nell'ultimo termine deve essere considerata su tutte le possibilità in cui si possono presentare gli altri $(N - k)$ eventi. Tale sommatoria vale quindi 1, da cui la tesi.

Capitolo 2

Distribuzioni discrete

2.1 Numeri aleatori con distribuzione discreta

X numero aleatorio si dice con *distribuzione discreta* se la cardinalità dell'insieme dei valori possibili $I(X)$ è finita o numerabile; la *distribuzione di probabilità* di X è data da

$$\mathbb{P}(X = x_i) = p(x_i) \qquad x_i \in I(X)$$

2.1.1 Schema di Bernoulli

Sia $(E_i)_{i \in \mathcal{N}}$ una successione di eventi stocasticamente indipendenti ed equiprobabili, ovvero tali che $\mathbb{P}(E_i) = p \; \forall i \in \mathcal{N}$, con $0 < p < 1$. Indipendenti vuol dire che, per ogni n, $E_1, \ldots, E_n$ sono stocasticamente indipendenti. Tale successione prende il nome di *schema di Bernoulli.*

Esempio 2.1 Un esempio di schema di Bernoulli è dato dalla successione di numeri aleatori che rappresentano il risultato dei lancio ripetuto di una moneta simmetrica.

2.1.2 Distribuzione binomiale

Dato $(E_i)_{i \in \mathcal{N}}$ uno schema di Bernoulli, sia S_n il numero aleatorio che conta i successi ottenuti su n prove. S_n si può scrivere come

$$S_n = E_1 + \ldots + E_n \; = \text{il numero di successi su } n \text{ prove}$$

L'insieme dei valori possibili per S_n è quindi $I(S_n) = \{0, \ldots, n\}$.
Calcoliamo, attraverso i costituenti, la distribuzione di probabilità di S_n.

$$\mathbb{P}(S_n = k) = \sum_{Q \subset (S_n = k)} \mathbb{P}(Q)$$

Per esempio, un costituente del primo tipo dell'evento $(S_n = k)$ è

$$Q = E_1 \cdots E_k \tilde{E}_{k+1} \cdots \tilde{E}_n$$

che rappresenta l'evento in cui i k successi si sono ottenuti con le prime k prove, mentre le restanti corrispondono ad insuccessi.

Analogamente, ogni altro costituente di $(S_n = k)$ conterrà k eventi che si sono verificati ed $(n-k)$ che non si sono verificati. Poiché gli E_i sono *iid*[1], si ottiene che ogni costituente Q ha la stessa probabilità, pari a

$$\mathbb{P}(Q) = \underbrace{p \cdots p}_{k \text{ volte}} \underbrace{(1-p) \cdots (1-p)}_{(n-k) \text{ volte}} = p^k (1-p)^{n-k}$$

Basta quindi contare quanti sono tali costituenti: essi sono $\binom{n}{k}$, pari al numero di modi di scegliere i k posti degli eventi che si verificano nella sequenza degli n eventi che compongono il costituente stesso. Si ottiene quindi

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Si dice che S_n ha *distribuzione binomiale* $\mathcal{B}(n,p)$ di parametri n, p.
Si verifica che $\sum_{k=0}^{n} \mathbb{P}(S_n = k) = 1$. Infatti, utilizzando le proprietà del binomio di Newton, si ottiene:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+1-p)^n = 1$$

Calcoliamo infine la previsione di X sapendo che $X = E_1 + \cdots + E_n$:

$$\mathbb{P}(X) = \mathbb{P}(E_1 + \cdots + E_n) = \sum_{i=1}^{n} \mathbb{P}(E_i) = np$$

Esempio 2.2 Consideriamo un'urna contenente N palline, di cui H bianche ed $N-H$ nere. Si fanno delle estrazioni con reimbussolamento. La successione $(E_i)_{i \in \mathcal{N}}$ di eventi $E_i =$ (si ottiene una pallina bianca all'i-esima estrazione) è uno schema di Bernoulli, mentre il numero aleatorio che conta il numero di palline bianche ottenute nelle prime n estrazioni ha distribuzione binomiale di parametri $\mathcal{B}(n, \frac{H}{N})$. Si veda l'esempio 1.11.

2.1.3 Distribuzione geometrica

Sia E_i uno schema di Bernoulli; sia T il numero aleatorio che rappresenta l'instante del primo successo in una serie di prove, ovvero $T = \inf\{n \mid E_n = 1\}$. L'insieme dei valori possibili per il numero aleatorio T è dato da:

$$I(T) = \mathcal{N} \setminus \{0\}$$

L'evento $(T = i)$ si può scrivere in termini degli E_i come

$$(T = i) = \tilde{E}_1 \cdots \tilde{E}_{i-1} E_i$$

Calcoliamo la distribuzione di probabilità:

$$\mathbb{P}(T = i) = \mathbb{P}\left(\tilde{E}_1 \cdots \tilde{E}_{i-1} E_i\right) = \mathbb{P}\left(\tilde{E}_1\right) \cdots \mathbb{P}\left(\tilde{E}_{i-1}\right) \mathbb{P}(E_i) = (1-p)^{i-1} p$$

[1]si indica con *iid* la proprietà di essere indipendenti e identicamente distribuiti.

Si dice che T ha *distribuzione geometrica* di parametro p. Utilizzando la somma della serie geometrica, si verifica che:

$$\sum_{i=1}^{+\infty} \mathbb{P}(T=i) = \sum_{i=1}^{+\infty}(1-p)^{i-1}p = p\sum_{k=0}^{+\infty}(1-p)^k = p\cdot\frac{1}{1-(1-p)} = 1$$

Si calcola la previsione di T utilizzando la formula

$$\mathbb{P}(T) = \sum_{i=1}^{+\infty} i\mathbb{P}(X=i) = \sum_{i=1}^{+\infty} i(1-p)^{i-1}p = p\sum_{i=1}^{+\infty} i(1-p)^{i-1} = p\frac{1}{p^2} = \frac{1}{p}$$

dove si è utilizzando il fatto che per la serie geometrica $\sum_{i=0}^{+\infty} x^i$ vale

$$\sum_{i=1}^{+\infty} ix^{i-1} = \sum_{i=1}^{+\infty} \frac{d}{dx}[x^i] = \frac{d}{dx}(\sum_{i=0}^{+\infty} x^i) = \frac{d}{dx}(\frac{1}{1-x}) = \frac{1}{(1-x)^2}$$

La distribuzione geometrica gode della proprietà di *"assenza di memoria"*. Vale infatti che

$$\mathbb{P}(T > m+n \mid T > n) = \mathbb{P}(T > m)$$

per ogni $m, n \in \mathcal{N}$. La proprietà di *assenza di memoria* ci dice che la probabilità di non avere un successo fino all'istante $m+n$ se non si era ancora ottenuto un successo fino all'istante n, è pari alla probabilità di non avere un successo fino all'istante m. Per dimostrare tale proprietà, basta osservare che

$$\mathbb{P}(T > m+n \mid T > n) = \frac{\mathbb{P}(T > m+n, T > n)}{\mathbb{P}(T > n)} = \frac{\mathbb{P}(T > m+n)}{\mathbb{P}(T > n)}$$

e che $\mathbb{P}(T > n) = (1-p)^n$ in quanto l'evento $(T > n)$ si verifica se e solo se i primi n eventi non si verificano. Ne segue allora che

$$\mathbb{P}(T > m+n \mid T > n) = \frac{\mathbb{P}(T > m+n)}{\mathbb{P}(T > n)} = \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = \mathbb{P}(T > m)$$

2.2 Distribuzione di Poisson

Un numero aleatorio X si dice avere distribuzione di Poisson di parametro λ, $\lambda \in \mathbb{R}_+$, se $I(X) = \mathcal{N}$ e vale che

$$\mathbb{P}(X=i) = \frac{\lambda^i}{i!}e^{-\lambda}$$

Si verifica che si ottiene una distribuzione di probabilità, ovvero che $\sum_{i=0}^{+\infty} \mathbb{P}(X=i) = 1$. Si ha che

$$\sum_{i=0}^{+\infty}\mathbb{P}(X=i)=\sum_{i=0}^{+\infty}\frac{\lambda^i}{i!}e^{-\lambda}=e^{-\lambda}\sum_{i=0}^{+\infty}\frac{\lambda^i}{i!}=e^{-\lambda}e^{\lambda}=1$$

Calcoliamo la previsione di X:

$$\mathbb{P}(X)=\sum_{i=0}^{+\infty}i\mathbb{P}(X=i)=\sum_{i=0}^{+\infty}i\frac{\lambda^i}{i!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{i=1}^{+\infty}\frac{\lambda^{i-1}}{(i-1)!}=\lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

2.3 La distribuzione ipergeometrica

Si consideri un'urna contentente N palline di cui H bianche ed $N-H$ nere. Si fanno n estrazioni senza reimbussolamento. Sia X il numero aleatorio che conta il numero di palline bianche nel campione.
Il minimo numero di palline bianche fra le n estratte sarà pari a 0 se nell'urna le palline nere sono più di n, mentre sarà dato da $n-(N-H)$ se una volta scelte tutte le palline nere rimarranno da determinare ancora degli elementi del campione. Viceversa, il numero massimo di palline bianche nel campione è dato dal minimo fra n ed il numero massimo di palline bianche nell'urna, ovvero H. Si ottiene che

$$I(X)=\{n-n\vee(N-H),\cdots,n\vee H\}$$

Sia $i\in I(X)$. Si vuole calcolare la distribuzione di probabilità di X utilizzando la formula

$$\mathbb{P}(X=i)=\frac{\sharp\, casi\ favorevoli}{\sharp\, casi\ possibili}$$

Il numero di casi possibili coincide con il numero di modi di scegliere n palline fra le N presenti nell'urna senza ripetizione e senza tener conto dell'ordine, ovvero

$$\sharp\, casi\ possibili=\binom{N}{n}$$

Per avere i palline bianche nel campione, bisogna prendere i palline bianche fra le H contenute nell'urna e scegliere le restanti $(n-i)$ fra le $(N-H)$ nere. Ne segue che

$$\sharp\, casi\ favorevoli=\binom{H}{i}\binom{N-H}{n-i}$$

Si dice che X possiede *distribuzione ipergeometrica* e vale che

$$\mathbb{P}(X=i)=\frac{\binom{H}{i}\binom{N-H}{n-i}}{\binom{N}{n}}$$

Sia E_i l'evento $E_i=$ (esce una pallina bianca alla i-esima estrazione). La probabilità di ottenere una pallina bianca alla i-esima estrazione è data da

$$\mathbb{P}(E_i) = \frac{\sharp\, casi\ favorevoli}{\sharp\, casi\ possibili} = \frac{HD_{n-1}^{N-1}}{D_n^N} = \frac{H}{N}$$

Infatti, se si considerano le n palline estratte come ordinate in una n-upla, il numero di casi favorevoli è dato dalle n-uple ordinate che hanno una pallina bianca all'i-esimo posto, mentre il numero dei casi possibili sono tutte le n-uple ordinate di n elementi scelti su N. Poiché $X = E_1 + \cdots + E_n$, usando la linerarità della previsione si ottiene

$$\mathbb{P}(X) = \sum_{i=1}^{n} \mathbb{P}(E_i) = n\frac{H}{N}$$

2.4 Indipendenza di partizioni

Si considerino due partizioni:

$$\mathcal{H} = (H_1, \ldots, H_m)$$
$$\mathcal{L} = (L_1, \ldots, L_n)$$

$\mathcal{H}$ e $\mathcal{L}$ si dicono *stocasticamente indipendenti* se per ogni i, j tali che $1 \leq i \leq m$, $1 \leq j \leq n$ vale

$$\mathbb{P}(H_i L_j) = \mathbb{P}(H_i)\,\mathbb{P}(L_j)$$

Date r partizioni $\mathcal{H}_1, \ldots, \mathcal{H}_r$, ciascuna formata da $n_i\,(i = 1, \ldots, r)$ eventi, esse si dicono *stocasticamente indipendenti* se per ogni scelta di indici $i_1, \ldots, i_r$ tali che $1 \leq i_1 \leq n_1, \ldots, 1 \leq i_r \leq n_r$ vale

$$\mathbb{P}\left(H_{i_1}^{(1)} \cdots H_{i_r}^{(r)}\right) = \mathbb{P}\left(H_{i_1}^{(1)}\right) \cdots \mathbb{P}\left(H_{i_r}^{(r)}\right)$$

dove $H_{i_k}^{(k)} \in \mathcal{H}_k$, $k = 1, \ldots, r$. Si può pensare ad una partizione come ad un plurievento.

2.5 Schema di Bernoulli generalizzato

Siano $\mathcal{H}_1, \ldots, \mathcal{H}_n$ partizioni contenenti lo stesso numero di eventi, $\mathcal{H}_i = \left\{E_1^{(i)}, \ldots, E_r^{(i)}\right\}$ $(i = 1, \ldots, n)$, tali che per ogni i valgano

1. $\forall j = 1, \ldots, r \quad \mathbb{P}\left(E_j^{(i)}\right) = p_j$

2. $p_1 + \cdots + p_r = 1$

Si suppone che $\mathcal{H}_1, \ldots, \mathcal{H}_n$ siano stocasticamente indipendenti. Si parla in questo caso di *schema di Bernoulli generalizzato*. La definizione si estende ad una successione infinita di partizioni $(\mathcal{H}_i)_{i \in \mathcal{N}}$ richiedendo che $\mathcal{H}_1, \ldots, \mathcal{H}_m$ soddisfino le condizioni predette per ogni m.

2.6 Distribuzione multinomiale

Consideriamo $Y_1, \ldots, Y_r$ numeri aleatori definiti come

$$Y_i = \sum_{k=1}^{n} E_i^{(k)}$$

Ne segue che

$$\sum_{i=1}^{r} Y_i = \sum_{i=1}^{r} \sum_{k=1}^{n} E_i^{(k)} = \sum_{k=1}^{n} \underbrace{\sum_{i=1}^{r} E_i^{(k)}}_{1} = n$$

Si calcola

$$\mathbb{P}\left(Y_1 = k_1, \ldots, Y_r = k_r\right) = \sum_{Q \text{ I tipo}} \mathbb{P}(Q) = \underbrace{\frac{n!}{k_1! \cdots k_r!}}_{\text{numero di costituenti}} \underbrace{p_1^{k_1} \cdots p_r^{k_r}}_{\mathbb{P}(Q)}$$

La *distribuzione multinomiale* dipende dal parametro r e dalle probabilità $p_1, \ldots, p_{r-1}$ (p_r è determinabile conoscendo le altre $r-1$ probabilità). Per $r = 2$ si ha la distribuzione binomiale.

2.7 Indipendenza stocastica per numeri aleatori

Siano X e Y due numeri aleatori con $I(X) = \{x_1, \ldots, x_m\}$ e $I(Y) = \{y_1, \ldots, y_n\}$. Considero le partizioni $\mathcal{H}$ generata dagli eventi $H_i = (X = x_i)$, $x_i \in I(X)$, e $\mathcal{L}$ generata dagli eventi $L_j = (Y = y_j)$, $y_j \in I(Y)$,.

X e Y si dicono *stocasticamente indipendenti* se lo sono le partizioni $\mathcal{H}$ e $\mathcal{L}$.

2.8 Distribuzione congiunta

Consideriamo il *vettore aleatorio* (X, Y) con insieme dei valori possibili $I(X, Y)$. Si definisce *distribuzione congiunta* di (X, Y) la probabilità

$$\mathbb{P}(X = x_i, Y = y_j)$$

dove $(x_i, y_j) \in I(X, Y)$. Si può associare alla distribuzione congiunta la matrice

$$\begin{pmatrix} p(x_1, y_1) & \cdots & p(x_1, y_n) \\ \vdots & \ddots & \vdots \\ p(x_m, y_1) & \cdots & p(x_m, y_n) \end{pmatrix}$$

dove $p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$. Si definisce *distribuzione marginale* di X

$$p_1(x_i) = \mathbb{P}(X = x_i)$$

Tale distribuzione marginale si ottiene dalla congiunta nel modo seguente:

$$p_1(x_i) = \mathbb{P}(X = x_i) = \sum_{j=1}^{n} \mathbb{P}(X = x_i | Y = y_j)\mathbb{P}(Y = y_j) = \sum_{j=1}^{n} \mathbb{P}(X = x_i, Y = y_j) = \sum_{j=1}^{n} p(x_i, y_j)$$

Analogamente, si definisce la *distribuzione marginale* di Y

$$p_2(y_j) = \mathbb{P}(Y = y_j) = \sum_{i=1}^{m} p(x_i, y_j)$$

Ne segue che due numeri aleatori sono stocasticamente indipendenti se e solo se

$$\forall (i,j) \quad p(x_i, y_j) = p_1(x_i)p_2(y_j) \tag{1}$$

Data $\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$, la previsione del numero aleatorio $Z = \psi(X,Y)$ si ottiene utilizzando la distribuzione congiunta di (X,Y) nel modo seguente:

$$\mathbb{P}(Z) = \mathbb{P}(\psi(X,Y)) = \sum_{(x_i,y_j)\in I(X,Y)} \psi(x_i, y_j)\mathbb{P}(X = x_i, Y = y_j)$$

La dimostrazione è analoga al caso unidimensionale.

2.9 Covarianza e varianza

Dati due numeri aleatori X e Y, si definisce *covarianza* di X e Y

$$\text{cov}(X,Y) = \mathbb{P}\left((X - \mathbb{P}(X))(Y - \mathbb{P}(Y))\right)$$

X e Y si dicono

- *correlati positivamente* se $\text{cov}(X,Y) > 0$
- *correlati negativamente* se $\text{cov}(X,Y) < 0$
- *non correlati* se $\text{cov}(X,Y) = 0$

Sviluppando la formula precedente, si ottiene

$$\text{cov}(X,Y) = \mathbb{P}(XY - \mathbb{P}(X)Y - X\mathbb{P}(Y) + \mathbb{P}(X)\mathbb{P}(Y)) = \mathbb{P}(XY) - \mathbb{P}(X)\mathbb{P}(Y)$$

La *varianza* è definita come

$$\sigma^2(X) = \text{cov}(X,X)$$

Si ottiene che $\sigma^2(X) = \mathbb{P}(X^2) - \mathbb{P}(X)^2$ ovvero $\sigma^2(X) = \mathbb{P}\left((X - \mathbb{P}(X))^2\right)$. Se $\sigma^2(X)$ è 0, allora X è una costante, ovvero tutta la probabilità è concentrata nella previsione $\mathbb{P}(X)$.

Data la varianza, si introducono inoltre

- *Scarto quadratico medio*:
$$\sigma(X) = \sqrt{\sigma^2(X)}$$

- *Previsione quadratica*:
$$P_Q(X) = \sqrt{\mathbb{P}(X^2)}$$

Proposizione 2.3 (Proprietà della covarianza e varianza) *La covarianza e la varianza rispettano le seguenti proprietà*

1. *La covarianza è bilineare, ovvero*
$$cov(X+Y,Z) = cov(X,Z) + cov(Y,Z)$$
2. $cov(aX+b, cY+d) = ac\ cov(X,Y)$
3. $\sigma^2(aX+b) = a^2\sigma^2(X)$

Dimostrazione.

1. Basta utilizzare il fatto che $\operatorname{cov}(X,Z) = \mathbb{P}(XZ) - \mathbb{P}(X)\mathbb{P}(Z)$ e la linerità della previsione. Si ottiene
$$\begin{aligned}\operatorname{cov}(X+Y,Z) &= \mathbb{P}[(X+Y)Z] - \mathbb{P}(X+Y)\mathbb{P}(Z)\\ &= \mathbb{P}(XZ+YZ) - [\mathbb{P}(X)+\mathbb{P}(Y)]\mathbb{P}(Z)\\ &= \mathbb{P}(XZ) - \mathbb{P}(X)\mathbb{P}(Z) + \mathbb{P}(YZ) - \mathbb{P}(Y)\mathbb{P}(Z)\\ &= \operatorname{cov}(X,Z) + \operatorname{cov}(Y,Z)\end{aligned}$$
2. Basta utilizzare la definizione di covarianza
$$\begin{aligned}\operatorname{cov}(aX+b, cY+d) &= \mathbb{P}\left((aX+b-\mathbb{P}(aX+b))\left(cY+d-\mathbb{P}(cY+d)\right)\right)\\ &= \mathbb{P}\left((aX+b-a\mathbb{P}(X)-b)\left(cY+d-c\mathbb{P}(Y)-d\right)\right)\\ &= \mathbb{P}\left(a\left(X-\mathbb{P}(X)\right)c\left(Y-\mathbb{P}(Y)\right)\right)\\ &= ac\operatorname{cov}(X,Y)\end{aligned}$$
3. Segue immediatamente dalla 1) sostituendo $(cY+d)$ con $(aX+b)$.

□

2.10 La varianza nelle distribuzioni discrete

Si calcola ora la varianza per le distribuzioni discrete viste in precedenza.

1. *La varianza nella somma di numeri aleatori*

   Proposizione 2.4 *Siano $X_1, \dots, X_n$ n numeri aleatori. Si ha che*

$$\sigma^2(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \sigma^2(X_i) + \sum_{\substack{i,j \\ i \neq j}} cov(X_i, X_j) =$$

$$= \sum_{i=1}^{n} \sigma^2(X_i) + 2 \sum_{i<j} cov(X_i, X_j)$$

   Dimostrazione.

$$\mathbb{P}\left(\left(\sum X_i - \sum \mathbb{P}(X_i)\right)^2\right) =$$

$$= \mathbb{P}\left[((X_1 - \mathbb{P}(X_1)) + \ldots + (X_n - \mathbb{P}(X_n)))^2\right] =$$

$$= \sum_{i=1}^{n} \underbrace{\mathbb{P}\left((X_i - \mathbb{P}(X_i))^2\right)}_{\sigma^2(X_i)} + \sum_{\substack{i,j \\ i \neq j}} \underbrace{\mathbb{P}\left((X_i - \mathbb{P}(X_i))(X_j - \mathbb{P}(X_j))\right)}_{\mathrm{cov}(X_i, X_j)}$$

   □

2. *Varianza di un evento*

$$\sigma^2(E_i) = \mathbb{P}\left(E_i^2\right) - \mathbb{P}\left(E_i\right)^2 = p(1-p)$$

3. *Distribuzione binomiale*: Si utilizza la rappresentazione $X = E_1 + \ldots + E_n$, dove gli E_i sono stocasticamente indipendenti. Si ottiene:

$$\sigma^2(E_1 + \ldots + E_n) = \sum_{i=1}^{n} \sigma^2(E_i) = np(1-p)$$

4. *Distribuzione geometrica*: sapendo che $\sigma^2(X) = \mathbb{P}[X^2] - \mathbb{P}(X)^2$, basta calcolare

$$\mathbb{P}(X^2) = \left(\sum_{i=1}^{+\infty} i^2 p(1-p)^{i-1}\right)$$

Si ottiene

$$\begin{aligned}
\mathbb{P}(X^2) = p\sum_{i=1}^{+\infty} i^2(1-p)^{i-1} &= \left(p\sum_{i=1}^{+\infty} i(i-1)(1-p)^{i-1}\right) + \mathbb{P}(X) \\
&= \left(p(1-p)\sum_{i=2}^{+\infty} i(i-1)(1-p)^{i-2}\right) + \mathbb{P}(X) \\
&= p(1-p)\frac{d}{dp}\left(-\sum i(1-p)^{i-1}\right) + \mathbb{P}(X) \\
&= p(1-p)\frac{d^2}{d^2p}\underbrace{\left(\sum (1-p)^i\right)}_{\frac{1}{p}} + \frac{1}{p} \\
&= \frac{2(1-p)}{p^2} + \frac{1}{p} \\
&= \frac{2}{p^2} - \frac{1}{p}
\end{aligned}$$

Infine

$$\begin{aligned}
\sigma^2(X) = \mathbb{P}[X^2] - \mathbb{P}(X)^2 &= \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{1}{p^2} - \frac{1}{p}
\end{aligned}$$

Per la distribuzione geometrica la varianza è data da

$$\sigma^2(X) = \frac{(1-p)}{p^2}$$

5. *Distribuzione di Poisson*: Se X ha distribuzione di Poisson di parametro λ, si calcola

$$P(X^2) = \sum_{i=0}^{+\infty} i^2\mathbb{P}(X=i) = \sum_{i=0}^{+\infty} i^2\frac{\lambda^i}{i!}e^{-\lambda} = e^{-\lambda}\sum_{i=0}^{+\infty}[(i^2-i)+i]\frac{\lambda^i}{i!} =$$

$$= \lambda^2 e^{-\lambda}\sum_{i=2}^{+\infty}\frac{\lambda^{i-2}}{(i-2)!} + \lambda e^{-\lambda}\sum_{i=1}^{+\infty}\frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} + \lambda = \lambda^2 + \lambda$$

Si ottiene che la varianza è data da

$$\sigma^2(X) = \mathbb{P}(X^2) - \mathbb{P}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

6. *Distribuzione ipergeometrica*: Nella stessa notazione della sezione 2.3, si utilizza la rappresentazione $X = E_1 + \ldots + E_n$. Tali eventi non sono stocasticamente indipendenti e risultano correlati negativamente. Infatti, se $H < N$ per ogni scelta di $i \neq j \in \{1, \ldots, n\}$ si ha

$$\mathrm{cov}(E_i, E_j) = \mathbb{P}(E_i E_j) - \mathbb{P}(E_i)\mathbb{P}(E_j) = \frac{H}{N^2}\frac{H-N}{N-1} < 0$$

in quanto

$$\mathbb{P}(E_i E_j) = \mathbb{P}(E_i \mid E_j)\mathbb{P}(E_j) == \frac{(H-1)D_{n-2}^{N-2}}{D_{n-1}^{N-1}}\frac{H}{N} = \frac{(H-1)}{N-1}\frac{H}{N}$$

La varianza di X si ottine utilizzando la formula della varianza della somma di n numeri aleatori

$$\sigma^2(X) = \sum_{i=1}^{n} \sigma^2(E_i) + \sum_{\substack{i,j \\ i \neq j}} \mathrm{cov}(E_i, E_j) = n\frac{H}{N}(1-\frac{H}{N}) + D_2^n \frac{H}{N^2}\frac{H-N}{N-1} = n\frac{N-n}{N-1}\frac{H}{N}(1-\frac{H}{N})$$

$D_2^n = \dfrac{n!}{(n-2)!}$ conta il numero numero di elementi nella sommatoria $\sum_{\substack{i,j \\ i \neq j}} \mathrm{cov}(E_i, E_j)$ che corrisponde al numero di coppie ordinate di elementi distinti scelti su n.

2.11 Il coefficiente di correlazione

Dati due numeri aleatori X, Y, si definisce *coefficiente di correlazione* di X, Y

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma(X)\,\sigma(Y)}$$

Le proprietà del coefficiente di correlazione sono:

1. $\rho(aX + b,\, cY + d) = \mathrm{segno}(ac)\,\rho(X, Y)$

   Utilizzando le proprietà della covarianza si ha che:

$$\begin{aligned} \rho(aX + b,\, cY + d) &= \frac{\mathrm{cov}(aX + b,\, cY + d)}{\sqrt{\sigma^2(aX + b)\,\sigma^2(cY + d)}} \\ &= \frac{ac\,\mathrm{cov}(X, Y)}{|ac|\,\sqrt{\sigma^2(X)\,\sigma^2(Y)}} \\ &= \mathrm{segno}(ac)\,\rho(X, Y) \end{aligned}$$

2. $-1 \leq \rho(X,Y) \leq 1$

Si considerano i numeri aleatori standardizzati

$$X^* = \frac{X - \mathbb{P}(X)}{\sigma(X)} \quad Y^* = \frac{Y - \mathbb{P}(Y)}{\sigma(Y)}$$

Dalla proprietà 2 della Proposizione 2.3, si ha che

$$\text{cov}(X^*, Y^*) = \mathbb{P}\left(X^* Y^*\right) = \frac{\mathbb{P}\left((X - \mathbb{P}(X))(Y - \mathbb{P}(Y))\right)}{\sigma(X)\,\sigma(Y)} = \rho(X,Y)$$

Calcolando la varianza di $X^* + Y^*$ si ottiene

$$\sigma^2(X^*+Y^*) = \sigma^2(X^*)+\sigma^2(Y^*)+2\,\text{cov}(X^*,Y^*) = 2+2\,\text{cov}(X^*,Y^*) = 2+2\rho(X,Y) \geq 0$$

Mentre dalla varianza di $X^* - Y^*$ segue che

$$\sigma^2(X^* - Y^*) = 2 - 2\rho(X,Y) \geq 0$$

Quindi vale

$$-1 \leq \rho(X,Y) \leq 1$$

2.12 Non correlazione ed indipendenza stocastica

Consideriamo due numeri aleatori X e Y con distribuzione congiunta discreta data da

$$p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$$

e distribuzioni marginali rispettivamente date da

$$p_i = \mathbb{P}(X = x_i) \qquad i = 1, \ldots, m$$

$$q_j = \mathbb{P}(Y = y_j) \qquad j = 1, \ldots, n$$

X e Y sono non correlati se e solo se

$$\mathbb{P}(XY) = \mathbb{P}(X)\mathbb{P}(Y)$$

ovvero se e solo se

$$\sum_{i,j} x_i y_j p_{i,j} = \sum_i x_i p_i \sum_j y_j q_j$$

Inoltre, devono valere le relazioni:

$$\begin{array}{ll} \sum_i p_i = 1 & \sum_j p_{i,j} = p_i \quad \forall i \\ \sum_j q_j = 1 & \sum_i p_{i,j} = q_j \quad \forall j \end{array}$$

$$\sum_{i,j} p_{i,j} = 1$$

Affinché X e Y risultino non correlati ed abbiano le p_i, q_j come distribuzioni marginali, su $mn-1$ parametri indipendenti bisogna imporre $1+(m-1)+(n-1)$ condizioni. Si ha $mn-1-(m-1)-(n-1)=(m-1)(n-1)-1$.

Le $m+n-1$ condizioni sono sufficienti a garantire la non correlazione, mentre l'indipendenza stocastica richiede la determinazione di $(mn-1)$ variabili. Ne segue che, per l'indipendenza stocastica, esistono delle soluzioni diverse da $p_{ij}=p_iq_j$. Segue immediatamente che la non correlazione non implica in generale l'indipendenza stocastica. Se $m=n=2$ allora il numero di relazioni è pari al numero di variabili; quindi, *due eventi sono non correlati se e solo se sono stocasticamente indipendenti.*
Si verifica che l'indipendenza stocastica implica invece la non correlazione. Se X, Y sono due numeri aleatori stocasticamente indipendenti con distribuzione discreta congiunta $p_{ij}=\mathbb{P}(X=x_i, Y=y_j)$, la loro covarianza $\operatorname{cov}(X,Y)$ si calcola nel modo seguente:

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \mathbb{P}(XY)-\mathbb{P}(X)\mathbb{P}(Y)\\
&= \sum_{i,j} x_i y_j p_{i,j} - \left(\sum_i x_i p_i\right)\left(\sum_j x_i q_j\right)\\
&= \sum_{i,j} x_i y_j p_i q_j - \left(\sum_i x_i p_i\right)\left(\sum_j x_i q_j\right)\\
&= 0
\end{aligned}$$

in quanto la distribuzione congiunta di due numeri aleatori stocasticamente indipendenti è data dal prodotto delle distribuzioni marginali p_i, q_j (proprietà (1)).

2.13 La disuguaglianza di Chebichev

Valgono le seguenti disuguaglianze, dette di Chebichev:

1. Sia X numero aleatorio tale che $P_Q(X)>0$. Per ogni $t>0$ vale che

$$\mathbb{P}\left(|X| \geq t P_Q(X)\right) \leq \frac{1}{t^2}$$

2. Sia X numero aleatorio con $\sigma^2(X)>0$. Posto $m=\mathbb{P}(X)$, per $\forall t>0$ si ha che:

$$\mathbb{P}\left(|X-m| \geq \sigma(X)t\right) \leq \frac{1}{t^2}$$

DIMOSTRAZIONE.

1. Sia E l'evento $E=(|X| \geq tP_Q(X))$. Calcoliamo $\mathbb{P}\left(X^2\right)$ con la formula delle probabilità totali:

$$\mathbb{P}\left(X^2\right)=\mathbb{P}\left(X^2|E\right)\mathbb{P}(E)+\mathbb{P}\left(X^2|\tilde{E}\right)\mathbb{P}\left(\tilde{E}\right)$$

Per la proprietà di monotonia della previsione, si ha che $\mathbb{P}\left(X^2|\tilde{E}\right) \geq 0$ in quanto X^2 è un numero aleatorio sempre positivo. Ne segue che

$$\mathbb{P}\left(X^2\right) \geq \mathbb{P}\left(X^2|E\right)\mathbb{P}(E) \geq t^2 P_Q(X)^2 \mathbb{P}(E)$$

Poiché $P_Q(X)^2 = \mathbb{P}(X^2)$ si ottiene

$$\mathbb{P}(E) \leq \frac{1}{t^2}$$

ovvero

$$\mathbb{P}\left(|X| \geq t P_Q(X)\right) \leq \frac{1}{t^2}$$

2. La seconda disuguaglianza segue dalla prima sostituendo ad X il numero aleatorio $Y = X - m$.

□

2.14 La legge debole dei grandi numeri

Teorema 2.5 *Sia $(X_n)_{n \in \mathcal{N}}$ una successione di numeri aleatori a due a due non correlati con stessa previsione $\mathbb{P}(X_i) = m$ e varianza $\sigma^2(X_i) = \sigma^2$.*

Posto $S_n = X_1 + \ldots + X_n$, si ha che

$$\forall \lambda > 0 \qquad \lim_{n \to +\infty} \mathbb{P}\left(|\frac{S_n}{n} - m| \geq \lambda\right) = 0$$

Il numero aleatorio $\frac{S_n}{n}$ si dice media campionaria.

Dimostrazione. Si dimostra il teorema utilizzando la seconda disuguaglianza di Chebichev. Si calcola la previsione di $\frac{S_n}{n}$

$$\mathbb{P}\left(\frac{S_n}{n}\right) = \frac{1}{n}\left(\mathbb{P}(X_1) + \ldots + \mathbb{P}(X_n)\right) = m$$

e la varianza di $\frac{S_n}{n}$

$$\sigma^2\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sigma^2(S_n) = \frac{1}{n^2}(\sum_{i=1}^{n}\sigma^2(X_i) + \sum_{i,j=1, i\neq j}^{n} \text{cov}(X_i, Y_j)) = \frac{\sigma^2}{n}$$

Dalla seconda disuguaglianza di Chebichev

$$\mathbb{P}\left(|\frac{S_n}{n} - m| \geq \frac{\sigma}{\sqrt{n}}t\right) \leq \frac{1}{t^2}$$

Posto $\lambda = \frac{\sigma}{\sqrt{n}}t$, si ricava $\frac{1}{t^2} = \frac{\sigma^2}{n\lambda^2}$. Ne segue che

$$\mathbb{P}\left(|\frac{S_n}{n} - m| \geq \lambda\right) \leq \frac{\sigma^2}{n\lambda^2} \longrightarrow_{n\to+\infty} 0$$

□

Esempio 2.6 $X_i = E_i$ eventi non correlati con $\mathbb{P}(E_i) = p$. Dalla legge dei grandi numeri segue che

$$\mathbb{P}\left(|\frac{E_1 + \ldots + E_n}{n} - p| \geq \lambda\right) \longrightarrow_{n\to+\infty} 0$$

In questo caso $\frac{S_n}{n} = \frac{E_1 + \ldots + E_n}{n}$ prende il nome di *frequenza*. Per un numero grande di prove, la frequenza approssima la probabilità di un evento.

Capitolo 3

Distribuzioni assolutamente continue uni-dimensionali

3.1 Funzione di ripartizione

Dato un numero aleatorio X, si introduce la *funzione di ripartizione* (o *distribuzione*) F di X data da

$$F : \mathbb{R} \to [0, 1]$$

$$F(x) = \mathbb{P}(X \leq x) \qquad x \in \mathbb{R}$$

Assegnare la distribuzione di probabilità di X significa specificare la sua funzione di ripartizione.

Esempio 3.1 (Caso discreto) Si può parlare di funzione di ripartizione anche nel caso di numeri aleatori con distribuzione discreta.

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i)$$

Se si conosce F, si può calcolare la probabilità degli intervalli

$$\mathbb{P}(a < X \leq b) = F(b) - F(a)$$

Proprietà

1. $0 \leq F(x) \leq 1$ (per definizione)
2. *monotonia*: $F(b) \geq F(a)$ se $b > a$ in quanto $F(b) - F(a) = \mathbb{P}(a < X \leq b) \geq 0$
3. *continuità a destra*: $F(x) = \lim_{y \to x^+} F(y)$
4. $\lim_{x \to +\infty} F(x) = 1$
5. $\lim_{x \to -\infty} F(x) = 0$

Le condizioni 3., 4., 5. sono proprietà aggiuntive di regolarità che saranno sempre verificate nei casi che considereremo. È possibile considerare casi in cui non valgono. Poiché F è monotona e limitata, il limite a sinistra esiste ed è finito. Nei casi che considereremo, tale limite è dato da:

$$\begin{aligned} F(x^-) &= \lim_{y \to x^-} F(y) \\ &= \lim_{y \to x^-} \mathbb{P}(X \leq y) \\ &= \mathbb{P}(X < x) \end{aligned}$$

Da cui $\mathbb{P}(X = x) = F(x) - F(x^-)$.

3.2 Distribuzioni assolutamente continue

Sia X un numero aleatorio con insieme dei valori possibili $I(X) = \mathbb{R}$, ovvero $I(X) = [a, b) \subset \mathbb{R}$[1]. Si dice che X ha *distribuzione assolutamente continua* se esiste $f : \mathbb{R} \to \mathbb{R}$ con le seguenti proprietá:

1. $\forall x \in \mathbb{R}, f(x) \geq 0$;
2. f sia *integrabile*;
3. $\int_{\mathbb{R}} f(s)\, ds = 1$.

tale che

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Tale funzione si dice *densità di probabilità*. Si noti che la f non è unica. Infatti, se cambiamo la f in un punto, la nuova funzione è ancora una densità di X.

Date la funzione di ripartizione e la funzione di densità di probabilità, vale la seguente uguaglianza:

$$f(x) = \frac{dF}{dx}$$

nei punti in cui f è continua.

Dall'ipotesi di regolarità 4. si ottiene

$$1 = \lim_{x \to +\infty} F(x) = \int_{-\infty}^{+\infty} f(s)\, \mathrm{d}s$$

Se f è continua nell'intervallo $[a, b]$, dal teorema del valor medio, esiste $\xi \in (a, b)$

[1] Eventualmente, l'intervallo può essere anche infinito.

$$\begin{aligned}\mathbb{P}(a < X \leq b) &= F(b) - F(a) \\ &= \int_a^b f(s)\,\mathrm{d}s \\ &= f(\xi)(b-a)\end{aligned}$$

da cui

$$f(\xi)(b-a) > 0 \Rightarrow f \geq 0$$

Si noti che la probabilità degli intervalli si calcola come:

$$\begin{aligned}\mathbb{P}(a < X \leq b) &= F(b) - F(a) \\ &= \int_{-\infty}^b f(s)\,\mathrm{d}s - \int_{-\infty}^a f(s)\,\mathrm{d}s \\ &= \int_a^b f(s)\,\mathrm{d}s\end{aligned}$$

Per quanto riguarda il calcolo della previsione nel caso assolutamente continuo, essa è data da

$$\mathbb{P}(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x$$

In generale, vale la formula

$$\mathbb{P}(\Psi(X)) = \int_{-\infty}^{+\infty} \Psi(x) f(x)\mathrm{d}x$$

dove $\Psi : \mathbb{R} \to \mathbb{R}$ è una funzione integrabile. Ne segue che la varianza si ottiene come

$$\begin{aligned}\sigma^2(X) &= \mathbb{P}(X^2) - \mathbb{P}(X)^2 \\ &= \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x - \left(\int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x\right)^2\end{aligned}$$

3.3 Distribuzioni assolutamente continue unidimensionali

3.3.1 Distribuzione uniforme in $[0,1]$

Un numero aleatorio X ha distribuzione uniforme in $[0,1]$ se la sua funzione di distribuzione è data da

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

La probabilità di un singolo punto è

$$\mathbb{P}(X = x) = F(x) - F(x^-) = 0$$

La densità di probabilità si definisce come

$$f(x) = F'(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x < 1 \\ 0 & x \geq 1 \end{cases}$$

Come nei casi seguenti, il valore della densità nei punti di discontinuità può essere scelto in maniera arbitraria.

3.3.2 Previsione

La previsione della distribuzione uniforme è

$$\mathbb{P}(X) = \int_{\mathbb{R}} x f(x)\,\mathrm{d}x = \int_0^1 x\,\mathrm{d}x = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

Varianza

La varianza è

$$\sigma^2(X) = \int_0^1 x^2\,\mathrm{d}x - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

3.3.3 Distribuzione uniforme su un intervallo qualunque $[a, b]$

Un numero aleatorio X ha distribuzione uniforme in $[a, b]$ se la sua funzione di distribuzione è data da

$$F(x) = \begin{cases} 0 & x \leq 0 \\ c(x-a) & a < x < b \\ 1 & x \geq 1 \end{cases}$$

Per calcolare la costante c:

$$F(b) = 1 \Rightarrow c(b-a) = 1 \Rightarrow c = \frac{1}{b-a}$$

Previsione

La previsione della distribuzione uniforme è

$$\mathbb{P}(X) = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{a+b}{2}$$

Varianza

La varianza è

$$\sigma^2(X) = \mathbb{P}((X - \mathbb{P}(X))^2) = \int_a^b \frac{1}{b-a}(x - \frac{a+b}{2})^2\,\mathrm{d}x = \frac{1}{b-a}\,\frac{1}{3}\left[\left(x - \frac{a+b}{2}\right)^3\right]_a^b = \frac{(b-a)^2}{12}$$

3.3.4 Distribuzione esponenziale di parametro λ

Un numero aleatorio X ha distribuzione esponenziale di parametro λ se la sua funzione di distribuzione è data da

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

La densità è data da:

$$f(x) = F'(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Se attibuiamo ad X il significato di un tempo aleatorio in cui si verifica un fatto (ad esempio l'istante di decadimento di un atomo), la distribuzione esponenziale ha la proprietà di *assenza di memoria*, ovvero dati $x, y \geq 0$

$$\mathbb{P}(X > x + y | X > y) = \mathbb{P}(X > x)$$

$$\begin{aligned} \mathbb{P}(X > x + y | X > y) &= \frac{\mathbb{P}(X > x + y, X > y)}{\mathbb{P}(X > y)} \\ &= \frac{\mathbb{P}(X > x + y)}{\mathbb{P}(X > y)} \\ &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} \\ &= e^{-\lambda x} \\ &= \mathbb{P}(X > x) \end{aligned}$$

Previsione

La previsione della distribuzione esponenziale è

$$\mathbb{P}(X) = \int_0^{+\infty} \lambda x e^{-\lambda x}\, \mathrm{d}x = \left[-x e^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\, \mathrm{d}x = \frac{1}{\lambda}$$

Varianza

La varianza della distribuzione esponenziale è

$$\begin{aligned} \sigma^2(X) &= \mathbb{P}(X^2) - \mathbb{P}(X)^2 \\ &= \int_0^{+\infty} \lambda x^2 e^{-\lambda x}\, \mathrm{d}x - \frac{1}{\lambda^2} \\ &= \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + 2\int_0^{+\infty} x e^{-\lambda x}\, \mathrm{d}x - \frac{1}{\lambda^2} \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2} \end{aligned}$$

3.3.5 Distribuzione normale

Un numero aleatorio X ha *distribuzione normale standard* (si indica con la notazione $N(0,1)$) se la sua funzione di densità è

$$f(x) = Ke^{-\frac{x^2}{2}} \quad , x \in \mathbb{R}$$

Calcolo della costante K

$$\begin{aligned}
\left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\,\mathrm{d}x\right)^2 &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\, e^{-\frac{y^2}{2}}\,\mathrm{d}x\mathrm{d}y \\
&= \int_{\mathbb{R}} e^{-\frac{x^2+y^2}{2}}\,\mathrm{d}x\mathrm{d}y \\
&= \int_0^{2\pi}\int_0^{+\infty} e^{-\frac{\rho^2}{2}}\,\rho\,\mathrm{d}\rho\mathrm{d}\theta \\
&= 2\pi\int_0^{+\infty} e^{-\frac{\rho^2}{2}}\,\rho\,\mathrm{d}\rho \\
&= 2\pi\left[-e^{-\frac{\rho^2}{2}}\right]_0^{+\infty} \\
&= 2\pi
\end{aligned}$$

dove si è effettuato il cambio di variabile $x = \rho\cos\theta$, $y = \rho\sin\theta$. Il determinante jacobiano di tale sostituzione è pari a ρ (so veda l'appendice B).

Ne segue che $K^{-2} = 2\pi$, ovvero $K^{-1} = \sqrt{2\pi}$, quindi

$$K = \frac{1}{\sqrt{2\pi}}$$

La funzione di ripartizione è

$$\mathcal{N}(x) = \int_{-\infty}^{x} n(t)\,\mathrm{d}t$$

dove $n(t) = f(t) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$.

Per la simmetria, si ottiene che

$$\mathcal{N}(-x) = 1 - \mathcal{N}(x)$$

Previsione

La previsione della distribuzione normale standard è

$$\mathbb{P}(X) = \int_{\mathbb{R}} x\; n(x)\,\mathrm{d}x = 0$$

poiché la funzione $f(x) = xe^{-\frac{x^2}{2}}$ è dispari.

Varianza

La varianza è

$$\sigma^2(X) = \int_{\mathbb{R}} \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x = \underbrace{\left[-\frac{x}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right]_{-\infty}^{+\infty}}_{\substack{\text{funzione dispari} \\ 0}} + \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x}_{\substack{\text{integrale della densità} \\ 1}} = 1$$

Introduciamo la distribuzione *normale*, indicata con la notazione $N(\mu, \sigma^2)$. Sia $X \sim N(0,1)$ e consideriamo $Y = \mu + \sigma X$, con $\sigma > 0$; la funzione di distribuzione di Y è

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(\mu + \sigma X \leq y) \\ &= \mathbb{P}\left(X \leq \frac{y-\mu}{\sigma}\right) \\ &= \mathcal{N}\left(\frac{y-\mu}{\sigma}\right) \end{aligned}$$

La densità di Y è allora

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\mathcal{N}\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma} n\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \end{aligned}$$

Stima delle code

Non esiste una formula in termini di funzioni elementari per la probabilità che $X \sim N(0,1)$ sia più grande di un $x > 0$ fissato. Possiamo darne delle stime asintotiche dall'alto e dal basso.

Proposizione 3.2 *Sia X un numero aleatorio con distribuzione normale standard. Per ogni $x > 0$, vale che*

$$\frac{n(x)}{x} - \frac{n(x)}{x^3} < \mathbb{P}(X > x) < \frac{n(x)}{x}$$

Il procedimento consiste nell'integrazione per parti della funzione di densità di probabilità $n(x)$.

prima integrazione per parti

$$\begin{aligned}\mathbb{P}(X > x) &= \int_x^{+\infty} n(t)\,\mathrm{d}t \\ &= \int_x^{+\infty} t\frac{n(t)}{t}\,\mathrm{d}t \\ &= \underbrace{\left[-\frac{n(t)}{t}\right]_x^{+\infty}}_{\frac{n(x)}{x}} - \int_x^{+\infty} \underbrace{\frac{n(t)}{t^2}}_{>0}\,\mathrm{d}t < \frac{n(x)}{x}\end{aligned}$$

seconda integrazione per parti

$$\begin{aligned}\mathbb{P}(X > x) &= \frac{n(x)}{x} - \int_x^{+\infty} t\frac{n(t)}{t^3}\,\mathrm{d}t \\ &= \frac{n(x)}{x} - \underbrace{\left[-\frac{n(t)}{t^3}\right]_x^{+\infty}}_{\frac{n(x)}{x^3}} + \int_x^{+\infty} \underbrace{\frac{3n(t)}{t^4}}_{>0}\,\mathrm{d}t > \frac{n(x)}{x} - \frac{n(x)}{x^3}\end{aligned}$$

3.3.6 Distribuzione gamma $\Gamma(\alpha, \lambda)$

Siano $\alpha, \lambda > 0$. Il numero aleatorio X ha *distribuzione gamma di parametri* α *e* λ se X è un numero aleatorio con distribuzione assolutamente continua di densità

$$g_{\alpha,\lambda}(x) = \begin{cases} K\,x^{\alpha-1}\,e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Si noti che la distribuzione esponenziale è un caso particolare di distribuzione gamma corrispondente alla scelta del parametro $\alpha = 1$.
Per calcolare la costante di normalizzazione K, si considera la *funzione gamma di Eulero* definita nel modo seguente:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}\,e^{-x}\,\mathrm{d}x$$

Proprietà di $\Gamma(\alpha)$

1. $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

   DIMOSTRAZIONE. Si procede integrando per parti.

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^{+\infty} x^{\alpha+1-1}\,e^{-x}\,\mathrm{d}x \\ &= \int_0^{+\infty} x^{\alpha}\,e^{-x}\,\mathrm{d}x \\ &= \left[-x^{\alpha}\,e^{-x}\right]_0^{+\infty} + \int_0^{+\infty} \alpha x^{\alpha-1}\,e^{-x}\,\mathrm{d}x \\ &= 0 + \alpha\int_0^{+\infty} x^{\alpha-1}\,e^{-x}\,\mathrm{d}x \\ &= \alpha\,\Gamma(\alpha)\end{aligned}$$

□

2. Se $\alpha = n$ allora $\Gamma(\alpha) = \Gamma(n) = (n-1)!$

Ne segue che la funzione Γ è un'estensione del fattoriale $n!$.

Calcolo della costante di normalizzazione K

Per calcolare il valore della costante di normalizzazione si procede imponendo che l'integrale della funzione di densità di probabilità sia uguale a 1.

$$\int_{-\infty}^{+\infty} g_{\alpha,\lambda}(s)\,\mathrm{d}s = \int_0^{+\infty} K\,x^{\alpha-1}\,e^{-\lambda x} = 1$$

Perciò si ottiene che

$$K = \frac{1}{\int_0^{+\infty} x^{\alpha-1}\,e^{-\lambda x}}$$

Calcoliamo tale integrale effettuando il cambio di variabile $y = \lambda x$

$$\begin{aligned} \int_0^{+\infty} x^{\alpha-1}\,e^{-\lambda x} &= \int_0^{+\infty} \frac{y^{\alpha-1}}{\lambda^{\alpha-1}}\,e^{-y}\,\frac{\mathrm{d}y}{\lambda} \\ &= \frac{1}{\lambda^\alpha}\int_0^{+\infty} y^{\alpha-1}\,e^{-y}\,\mathrm{d}y \\ &= \frac{\Gamma(\alpha)}{\lambda^\alpha} \end{aligned}$$

Ne segue che la costante di normalizzazione c è data da

$$K = \frac{\lambda^\alpha}{\Gamma(\alpha)}$$

Previsione di $\Gamma(\alpha, \lambda)$

$$\begin{aligned} \mathbb{P}(X) &= \int_{\mathbb{R}} x g_{\alpha,\lambda}(x)\,\mathrm{d}x \\ &= \int_0^{+\infty} x\,\frac{\lambda^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}\,e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^{+\infty} x^\alpha\,e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\,\frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\,\frac{\alpha\,\Gamma(\alpha)}{\lambda^{\alpha+1}} \\ &= \frac{\alpha}{\lambda} \end{aligned}$$

la previsione è quindi data da

$$\mathbb{P}(X) = \frac{\alpha}{\lambda}$$

Varianza di $\Gamma(\alpha, \lambda)$

$$\sigma^2(X) = \mathbb{P}(X^2) - \frac{\alpha^2}{\lambda^2}$$

$$\begin{aligned}
\mathbb{P}(X^2) &= \int_0^{+\infty} x^2 \, \frac{\lambda^\alpha}{\Gamma(\alpha)} \, x^{\alpha-1} \, e^{-\lambda x} \, \mathrm{d}x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha+1} \, e^{-\lambda x} \, \mathrm{d}x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \, \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \, \frac{(\alpha+1) \, \alpha \, \Gamma(\alpha)}{\lambda^{\alpha+2}} \\
&= \frac{\alpha \, (\alpha+1)}{\lambda^2}
\end{aligned}$$

Dalla precedente si ottiene

$$\sigma^2(X) = \frac{\alpha \, (\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

Osservazione sulla distribuzione Γ

La distribuzione esponenziale è un caso particolare di distribuzione Γ che si ottiene ponendo $\alpha = 1$, ovvero la distribuzione esponenziale è espressa dal caso particolare $\Gamma(1, \lambda)$.

Capitolo 4

Distribuzioni assolutamente continue n-dimensionali

4.1 Distribuzioni bidimensionali

Consideriamo il vettore aleatorio (X, Y). La funzione di ripartizione congiunta di (X, Y) è definita

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

La funzione di ripartizione è quindi una funzione da $\mathbb{R}^2$ in $[0, 1]$.

$$F : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

La probabilità degli intervalli si calcola come

$$\begin{aligned} \mathbb{P}(a_1 < X \leq b_1, a_2 < Y \leq b_2) = \mathbb{P}\left[((X \leq b_1)\backslash(X \leq a_1))\,((Y \leq b_2)\backslash(Y \leq a_2))\right] &= \\ \mathbb{P}(X \leq b_1, Y \leq b_2) - \mathbb{P}(X \leq a_1, Y \leq b_2) - \mathbb{P}(X \leq b_1, Y \leq a_2) + \mathbb{P}(X \leq a_1, Y \leq a_2) &= \\ F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) & \end{aligned}$$

Proprietà

1. $0 \leq F(x, y) \leq 1$

2. *Ipotesi di continuità* (che supporremo sempre verificate):

$$(i) \qquad \lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$$

$$(ii) \qquad \lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0$$

$$(iii) \qquad \lim_{\substack{x \to x_0^+ \\ y \to y_0^+}} F(x, y) = F(x_0, y_0)$$

In queste ipotesi, si ottiene che

$$\mathbb{P}(X = x_0, Y = y_0) = F(x_0, y_0) - F(x_0^-, y_0) - F(x_0, y_0^-) + F(x_0^-, y_0^-)$$

4.1.1 Funzioni di ripartizione marginali

Dati X, Y due numeri aleatori con funzione di ripartizione congiunta $F(x, y)$, si dicono *funzioni di ripartizione marginali* le funzioni di ripartizione F_1, F_2 di X e Y.
La funzione di ripartizione marginale per il numero aleatorio X si determina nel modo seguente

$$F_1(x) = \mathbb{P}(X \leq x) = \lim_{y \to +\infty} F(x, y)$$

Analogamente, si determina la funzione di ripartizione marginale per Y

$$F_2(y) = \mathbb{P}(Y \leq y) = \lim_{x \to +\infty} F(x, y)$$

Due numeri aleatori si dicono *stocasticamente indipendenti* se:

$$F(x, y) = F_1(x) F_2(y)$$

per ogni coppia $(x, y) \in \mathbb{R}^2$.

Proposizione 4.1 *Due numeri aleatori X, Y sono stocasticamente indipendenti se e solo per ogni quadrupla (a, b, c, d) di numeri reali vale che*

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \mathbb{P}(a < X \leq b)\mathbb{P}(c < Y \leq d)$$

DIMOSTRAZIONE.

$$\begin{aligned}
\mathbb{P}(a < X \leq b, c < Y \leq d) &= F(b, d) - F(b, c) - F(a, d) + F(a, c) \\
&= F_1(b)F_2(d) - F_1(b)F_2(c) - F_1(a)F_2(d) + F_1(a)F_2(c) \\
&= (F_1(b) - F_1(a))(F_2(d) - F_2(c)) \\
&= \mathbb{P}(a < X \leq b)\mathbb{P}(c < Y \leq d)
\end{aligned}$$

□

4.1.2 Caso assolutamente continuo

Il vettore aleatorio (X, Y) ha distribuzione assolutamente continua se esiste

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

tale che

1. f sia *non negativa* e *integrabile*
2. $\int\int_{\mathbb{R}^2} f(x, y)\, \mathrm{d}x\mathrm{d}y = 1$

e vale che

$$F(x,y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,\mathrm{d}s\mathrm{d}t$$

Tale $f(x,y)$ si dice *densità congiunta*. Applicando la formula per la probabilità degli intervalli, si ottiene

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \int_{a}^{b} \int_{c}^{d} f(s,t)\,\mathrm{d}s\mathrm{d}t$$

In generale, la probabilità che il vettore aleatorio (X,Y) appartenga ad una regione A del piano $\mathbb{R}^2$ è data dall'integrale della densità congiunta su A

$$\mathbb{P}((X,Y) \in A) = \int\int_{A} f(s,t)\,\mathrm{d}s\mathrm{d}t$$

Inoltre, se $\psi : \mathbb{R}^2 \to \mathbb{R}$ è una funzione sufficientemente regolare, posto $Z = \psi(X,Y)$, come nel caso unidimensionale si ha che

$$\mathbb{P}(Z) = \int\int_{\mathbb{R}^2} \psi(s,t) f(s,t)\,\mathrm{d}s\mathrm{d}t$$

Per esempio, se $Z = XY$ si ottiene che

$$\mathbb{P}(XY) = \int\int_{\mathbb{R}^2} st f(s,t)\,\mathrm{d}s\mathrm{d}t$$

Per calcolare le *densità marginali* di X e Y si usa il fatto che la funzione di ripartizione marginale si calcola come

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{x} f(s,t)\,\mathrm{d}s\mathrm{d}t = \int_{-\infty}^{x} \left(\int_{\mathbb{R}} f(s,t)\,\mathrm{d}t \right) \mathrm{d}s$$

Ne segue che la densià marginale è data da

$$f_X(x) = \int_{-\infty}^{+\infty} f(s,t)\,\mathrm{d}t$$

Analogamente,

$$f_Y(y) = \int_{-\infty}^{+\infty} f(s,t)\,\mathrm{d}s$$

Dalla definizione di indipendenza stocastica, segue che X ed Y sono stocasticamente indipendenti *se e solo se*

$$f(x,y) = f_X(x) f_Y(y)$$

ovvero la densità congiunta è uguale al prodotto delle densità marginali.

4.1.3 La densità di $Z = X + Y$

Siano X ed Y due numeri aleatori con densità congiunta $f(x, y)$. Si vuole calcolare la densità di

$$Z = X + Y$$

Si calcola la funzione di ripartizione

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) \\ &= \mathbb{P}(X + Y \leq z) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x, y) \, \mathrm{d}y \mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \mathrm{d}x \int_{-\infty}^{z} f(x, t - x) \, \mathrm{d}t \\ &= \int_{-\infty}^{z} \mathrm{d}t \int_{-\infty}^{+\infty} f(x, t - x) \, \mathrm{d}x \end{aligned}$$

dove si è effettuato il cambio coordinate $x = x$ e $t = x+y$ con corrispondente determinante jacobiano $|\det J| = 1$. Derivando si ottiene che la densità di Z è

$$f_Z(z) = F_Z'(z) = \int_{\mathbb{R}} f(x, z - x) \, \mathrm{d}x$$

In particolare, se $f(x, y) = f_X(x) f_Y(y)$, allora

$$f_Z(z) = \int_{\mathbb{R}} f_X(x) f_Y(z - x) \, \mathrm{d}x$$

Come esempio particolare di questa formula, si considerino due numeri aleatori X ed Y stocasticamente indipendenti e, rispettivamente, di densità $\Gamma(\alpha, \lambda)$ e $\Gamma(\beta, \lambda)$ (ovvero densità gamma di parametri α, λ e β, λ). La densità congiunta di (X, Y) è data da

$$f(x, y) = f_X(x) f_Y(y).$$

Utilizzando la formula precedente, si calcola la densità di $Z = X + Y$.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f(x, z - x) \, \mathrm{d}x \\ &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) \, \mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} \, x^{\alpha-1} \, e^{-\lambda x} \, I_{\{x>0\}} \, \frac{\lambda^\beta}{\Gamma(\beta)} \, (z - x)^{\beta-1} \, e^{-\lambda(z-x)} \, I_{\{(z-x)>0\}} \, \mathrm{d}x \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \, e^{-\lambda z} \int_{-\infty}^{+\infty} x^{\alpha-1} (z - x)^{\beta-1} I_{\{x>0,(z-x)>0\}} \, \mathrm{d}x \end{aligned}$$

Dato $A \subset \mathbb{R}^n$, si ricordi che la funzione $I_A(x)$ si chiama *funzione indicatrice* ed è definita nel seguente modo

$$I_A(x) = \begin{cases} 0 & x \notin A \\ \\ 1 & x \in A \end{cases}$$

Si ottiene dunque

- Se $z \leq 0$, allora $f_Z(z) = 0$;
- Se $z > 0$ allora $0 < x < z$ e

$$\begin{aligned} f_Z(z) &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1}(z-x)^{\beta-1}\,\mathrm{d}x \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^1 (zt)^{\alpha-1}(z-zt)^{\beta-1} z\,\mathrm{d}t \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha+\beta-1} e^{-\lambda z} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t\, z^{\alpha+\beta-1} e^{-\lambda z} \\ &= K\, z^{\alpha+\beta-1} e^{-\lambda z} \end{aligned}$$

dove si é effettuato il cambio di variabili $x = zt$, $\mathrm{d}x = z\mathrm{d}t$ e si è posto

$$K = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\,\Gamma(\beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$$

Ne segue che Z ha distribuzione $\Gamma(\alpha+\beta, \lambda)$. Quindi, se X e Y sono stocasticamente indipendenti e hanno entrambi distribuzione esponenziale di parametro λ, allora

$$Z = X + Y \sim \Gamma(1,\lambda) + \Gamma(1,\lambda) = \Gamma(2,\lambda)$$

Si ha inoltre

$$K = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)}$$

$$\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\,\Gamma(\beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)}$$

$$\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

ovvero si è ottenuto il valore dell'integrale $\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$ in termini della funzione Γ.

Si può introdurre la seguente distribuzione assolutamente continua importante in statistica.

4.1.4 La distribuzione beta $\mathcal{B}(\alpha, \beta)$

Siano $\alpha, \beta > 0$. Un numero aleatorio X ha distribuzione beta $\mathcal{B}(\alpha,\beta)$ se ha densità

$$f(x) = \begin{cases} K\, x^{\alpha-1}(1-x)^{\beta-1} & x \in [0,1] \\ 0 & \textit{altrimenti} \end{cases}$$

con $\alpha, \beta > 0$. Dai conti precedenti, si ottiene immediatamente

$$K = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} = \frac{1}{\int_0^1 x^{\alpha-1}\,(1-x)^{\beta-1}\,\mathrm{d}x}$$

La previsione

$$\begin{aligned}
\mathbb{P}(X) &= \int_0^1 x\,f(x)\,\mathrm{d}x \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,x^{\alpha}\,(1-x)^{\beta-1}\,\mathrm{d}x \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\Gamma(\alpha+1)\,\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\alpha\,\Gamma(\alpha)\,\Gamma(\beta)}{(\alpha+\beta)\,\Gamma(\alpha+\beta)} \\
&= \frac{\alpha}{(\alpha+\beta)}
\end{aligned}$$

La varianza

$$\sigma^2(X) = \mathbb{P}(X^2) - \mathbb{P}(X)^2 = \mathbb{P}(X^2) - \frac{\alpha^2}{(\alpha+\beta)^2}$$

Calcolo $\mathbb{P}(X^2)$

$$\begin{aligned}
\mathbb{P}(X^2) &= \int_0^1 x^2\,f(x)\,\mathrm{d}x \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,x^{\alpha+1}\,(1-x)^{\beta-1}\,\mathrm{d}x \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\Gamma(\alpha+2)\,\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{(\alpha+1)\,\alpha\,\Gamma(\alpha)\,\Gamma(\beta)}{(\alpha+\beta+1)\,(\alpha+\beta)\,\Gamma(\alpha+\beta)} \\
&= \frac{(\alpha+1)\,\alpha}{(\alpha+\beta+1)\,(\alpha+\beta)}
\end{aligned}$$

Si ottiene

$$\begin{aligned}\sigma^2(X) &= \mathbb{P}(X^2) - \mathbb{P}(X)^2 \\ &= \frac{(\alpha+1)\,\alpha}{(\alpha+\beta+1)\,(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\ &= \frac{\alpha}{(\alpha+\beta)}\,\frac{(\alpha+1)(\alpha+\beta) - \alpha\,(\alpha+\beta+1)}{(\alpha+\beta)\,(\alpha+\beta+1)} \\ &= \frac{\alpha}{(\alpha+\beta)}\,\frac{\alpha^2+\alpha+\alpha\beta+\beta-\alpha^2-\alpha-\alpha\beta}{(\alpha+\beta)\,(\alpha+\beta+1)} \\ &= \frac{\alpha\,\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}\end{aligned}$$

4.2 Distribuzioni n-dimensionali

Sia $(X_1, X_2, \ldots, X_n)$ un vettore aleatorio di dimensione n. La funzione

$$F : \mathbb{R}^n \longrightarrow [0,1]$$

definita come

$$F(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

si dice *funzione di ripartizione* di $(X_1, X_2, \ldots, X_n)$.

Proprietà

1. $\lim_{\substack{x_1 \to +\infty \\ \vdots \\ x_n \to +\infty}} F(x_1, x_2, \ldots, x_n) = 1$

2. $\lim_{x_i \to -\infty} F(x_1, x_2, \ldots, x_n) = 0$

4.2.1 Distribuzioni assolutamente continue n-dimensionali

Il vettore aleatorio $(X_1, X_2, \ldots, X_n)$ ha distribuzione assolutamente continua se esiste una funzione

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

tale che

1. f sia *non negativa* e *integrabile*

2. $\int_{\mathbb{R}^n} f(x_1, \cdots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n = 1$

e vale che

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, t_2, \ldots, t_n)\, \mathrm{d}t_1 \mathrm{d}t_2 \ldots \mathrm{d}t_n$$

Le funzioni di densità marginale si calcolano nel seguente modo

$$f_{X_{i_1},\dots,X_{i_r}}(x_{i_1},\dots,x_{i_r}) = \int_{\mathbb{R}^{n-r}} f(x_{i_1},\dots,x_{i_r},t_{i_{r+1}},\dots,t_{i_n})\,\mathrm{d}t_{i_{r+1}}\dots\mathrm{d}t_{i_n}$$

per ogni scelta di indici $i_1,\dots,i_r$ in $\{1,\dots,n\}$.

4.2.2 Distribuzione gaussiana n-dimensionale

Un vettore aleatorio $(X_1, X_2, \dots, X_n)$ ha *distribuzione gaussiana* n-dimensionale se ha densità

$$f(x_1, x_2, \dots, x_n) = Ke^{-\frac{1}{2}Ax\cdot x + b\cdot x}$$

dove $x = (x_1, x_2, \dots, x_n)^t \in \mathbb{R}^n$, A è una matrice

- *simmetrica*: $A^t = A$, ovvero $a_{ij} = a_{ji}$
- *definita positiva*: $Ax \cdot x \geq 0, \forall x \in \mathbb{R}^n$, e $Ax \cdot x = 0$ implica che $x = 0$

e $b = (b_1, b_2, \dots, b_n)^t$ è un vettore in $\mathbb{R}^n$. Il simbolo A^t indica la *matrice trasposta*, ovvero di elementi

$$[A^t]_{i,j} = [A]_{j,i}$$

Si ricordi inoltre che $b \cdot x$ indica il *prodotto scalare* fra il vettore b ed il vettore x, ovvero

$$b \cdot x = \sum_{i=1}^{n} b_i x_i$$

mentre Ax è il vettore che si ottiene come prodotto della matrice A per il vettore x, le cui componenti sono

$$[Ax]_i = \sum_j a_{ij} x_j$$

L'espressione $Ax \cdot x$ è una forma quadratica del tipo

$$Ax \cdot x = \sum_{i,j} a_{ij} x_i x_j$$

Viceversa, se si parte da una forma quadratica

$$\sum_{i,j} \alpha_{ij} x_i x_j$$

ci si può sempre ricondurre ad una rappresentazione matriciale associata ad una matrice simmetrica di componenti:

$$a_{ij} = \begin{cases} \alpha_{ii} & i = j \\ (\alpha_{ij} + \alpha_{ji})/2 & i \neq j \end{cases}$$

Caso 1: A diagonale e $b = 0$

Siano

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

e $b = 0$. Si ottiene

$$f(x_1, x_2, \ldots, x_n) = K \exp - \left(\lambda_1 \frac{x_1^2}{2} + \lambda_2 \frac{x_2^2}{2} + \ldots + \lambda_n \frac{x_n^2}{2} \right)$$

ovvero

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

dove

$$f_{X_i}(x_i) = \sqrt{\frac{\lambda_i}{2\pi}}\, e^{-\frac{\lambda_i x_i^2}{2}}$$

è la densità marginale di X_i. Ne segue che:

1. $X_1, \ldots, X_n$ sono stocasticamente indipendenti

2. Ogni X_i ha densità gaussiana $N\left(0, \frac{1}{\lambda_i}\right)$

3. La costante di normalizzazione è data da

$$\begin{aligned} K &= \sqrt{\frac{\lambda_1}{2\pi}} \sqrt{\frac{\lambda_2}{2\pi}} \cdots \sqrt{\frac{\lambda_n}{2\pi}} \\ &= \sqrt{\frac{\lambda_1 \lambda_2 \cdots \lambda_n}{(2\pi)^n}} \\ &= \sqrt{\frac{\det A}{(2\pi)^n}} \end{aligned}$$

Il *vettore delle previsioni* di $(X_1, X_2, \ldots, X_n)$ è allora

$$(\mathbb{P}(X_1), \mathbb{P}(X_2), \ldots, \mathbb{P}(X_n)) = (0, 0, \ldots, 0)$$

e la *matrice di covarianza*

$$C = \begin{pmatrix} \sigma^2(X_1) & \mathrm{cov}(X_1,X_2) & \cdots & \mathrm{cov}(X_1,X_n) \\ \mathrm{cov}(X_2,X_1) & \sigma^2(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{cov}(X_{n-1},X_n) \\ \mathrm{cov}(X_n,X_1) & \cdots & \mathrm{cov}(X_n,X_{n-1}) & \sigma^2(X_n) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\lambda_n} \end{pmatrix}$$

$$= A^{-1}$$

Caso 2: A diagonale e $b \neq 0$

Per ricondursi al caso $b = 0$ si utilizza la *traslazione* $X = U + c$ di componenti $[X]_i = [U]_i + [c]_i$, dove c è un vettore di $\mathbb{R}^n$. Si ottiene

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f(u_1 + c_1, u_2 + c_2, \ldots, u_n + c_n) \\ &= K' \exp\,[-\frac{1}{2}A(u+c)\cdot(u+c) + b\cdot(u+c)] \\ &= K' \exp\,[-\frac{1}{2}Au\cdot u - \frac{1}{2}Au\cdot c - \frac{1}{2}Ac\cdot u - \frac{1}{2}Ac\cdot c + b\cdot u + b\cdot c] \\ &= \underbrace{K' \exp\,[-\frac{1}{2}Ac\cdot c + b\cdot c]}_{\text{costante}} \exp\,[-\frac{1}{2}Au\cdot u + (b - Ac)\cdot u] \end{aligned}$$

in quanto

$$Ac\cdot u = Au\cdot c$$

perché A è simmetrica. Per ricondursi al caso precedente, bisogna annullare la parte di primo grado in U, quindi si sceglie

$$b - Ac = 0 \quad \Rightarrow \quad c = A^{-1}b$$

(A è invertibile in quanto definita positiva). Per tale scelta di c

$$\begin{aligned} f(u_1+c_1,u_2+c_2,\dots,u_n+c_n) &= K' \exp[A^{-1}b\cdot b - \frac{A(A^{-1}b)\cdot A^{-1}b}{2}]\exp[-\frac{1}{2}Au\cdot u] \\ &= \underbrace{K' \exp[\frac{1}{2}A^{-1}b\cdot b]}_{K''}\exp[-\frac{1}{2}Au\cdot u] \\ &= K'' \exp[-\frac{1}{2}Au\cdot u] \end{aligned}$$

Usando i risultati precedenti, si ottiene

1. la previsione $\mathbb{P}(X_i) = \mathbb{P}(U_i + c_i) = P(U_i) + c_i = 0 + c_i \quad \Rightarrow \quad \mathbb{P}(X_i) = (A^{-1}b)_i$, ovvero in notazione vettoriale:

$$\mathbb{P}(X) = A^{-1}b$$

2. La costante di normalizzazione è

$$K' = \sqrt{\frac{\det A}{(2\pi)^n}}\, e^{-\frac{1}{2}A^{-1}b\cdot b}$$

in quanto $K'' = \sqrt{\frac{\det A}{(2\pi)^n}}$.

3. La matrice di covarianza di X è la stessa di U in quanto una traslazione lascia invariate le covarianze.

$$C = \begin{pmatrix} \sigma^2(X_1) & \mathrm{cov}(X_1,X_2) & \cdots & \mathrm{cov}(X_1,X_n) \\ \mathrm{cov}(X_2,X_1) & \sigma^2(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{cov}(X_{n-1},X_n) \\ \mathrm{cov}(X_n,X_1) & \cdots & \mathrm{cov}(X_n,X_{n-1}) & \sigma^2(X_n) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2(U_1) & \mathrm{cov}(U_1,U_2) & \cdots & \mathrm{cov}(U_1,U_n) \\ \mathrm{cov}(U_2,U_1) & \sigma^2(U_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{cov}(U_{n-1},U_n) \\ \mathrm{cov}(U_n,U_1) & \cdots & \mathrm{cov}(U_n,U_{n-1}) & \sigma^2(U_n) \end{pmatrix}$$

Ne segue che

$$C = A^{-1}$$

Caso 3: A non diagonale e $b \neq 0$

Se A non è diagonale, le X_i non sono più stocasticamente indipendenti perciò la densità congiunta non è data dal prodotto delle densità marginali. A meno di fare una traslazione, si può pensare di avere $b = 0$. Poiché A è simmetrica, esiste una trasformazione ortogonale O che diagonalizza A, ovvero tale $O^t A O = D$, dove D è diagonale e $O^t = O^{-1}$.

Se si considera la trasformazione $X = OU$ si ottiene

$$
\begin{aligned}
f(u_1, \ldots, u_n) &= K \exp\left[-\frac{1}{2} A\, Ou \cdot Ou\right] \\
&= K \exp\left[-\frac{1}{2} O^t A O u \cdot u\right] \\
&= K \exp\left[-\frac{1}{2} Du \cdot u\right]
\end{aligned}
$$

Ci siamo ricondotti al primo caso, in cui si aveva A diagonale e $b = 0$. La matrice di covarianza di X, in notazione multidimensionale, è data da

$$
\begin{aligned}
C &= \mathbb{P}(X X^t) \\
&= \mathbb{P}(OU\,(OU)^t) \\
&= O^t\, \mathbb{P}(U\, U^t)\, O \\
&= O^t\, D^{-1}\, O \\
&= A^{-1}
\end{aligned}
$$

Si vede facilmente che, se Z è una matrice aleatoria ed A, B sono matrici costanti, si ha che $\mathbb{P}(AZB) = A\mathbb{P}(Z)B$ (ovviamente si suppone che sia possibile effettuare il prodotto). Riassumendo nel caso in A sia non diagonale e $b \neq 0$, si ottiene:

1. *Costante di normalizzazione*

$$K = \sqrt{\frac{\det A}{(2\pi)^n}}\, e^{-\frac{1}{2} A^{-1} b \cdot b}$$

2. *previsione*

$$\mathbb{P}(X) = A^{-1} b$$

3. *matrice di varianza-covarianza*

$$C = A^{-1}$$

Nota

Anche in questo caso, le distribuzioni marginali delle X_i sono gaussiane $N(\mathbb{P}(X_i), \sigma^2(X_i))$. Per la distribuzione gaussiana, $\mathrm{cov}(X_i, X_j) = 0$ implica l'indipendenza stocastica di X_i e X_j.

Caso particolare: $n = 2$

La matrice di varianza-covarianza è data da:

$$C = \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

dove con ρ si indica il coefficiente di correlazione. Si può quindi ricavare la matrice A dalla matrice di covarianza C nel seguente modo:

$$\begin{aligned} A = C^{-1} &= \frac{1}{\det C} \begin{pmatrix} \sigma_2^2 & -\rho\,\sigma_1\sigma_2 \\ -\rho\,\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \\ &= \frac{1}{\sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2} \begin{pmatrix} \sigma_2^2 & -\rho\,\sigma_1\sigma_2 \\ -\rho\,\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \\ &= \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \end{aligned}$$

Quindi, la densità gaussiana bidimensionale con parametri $m_1, m_2, \sigma_1, \sigma_2, \rho$ è data da:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_1)^2}{\sigma_1^2} - 2\frac{\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)\right]$$

Capitolo 5

Catene di Markov a tempo discreto

5.1 Catene di Markov omogenee con un numero finito di stati

Definiamo una *catena di Markov* con *spazio degli stati* S, dove S è un insieme finito, come Sia $(X_i)_{i\in\mathcal{N}}$ una successione di numeri aleatori $(X_i)_{i\in\mathcal{N}}$ con $I(X_i)_{i\in\mathcal{N}} = S$ tale che

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) = \rho_{s_0} p_{s_0,s_1} p_{s_1,s_2} \cdots p_{s_{n-1},s_n}$$

dove

1. ρ_{s_i}, $i = 1, \ldots, n$, è detta la *distribuzione iniziale*
$$\rho_{s_i} = \mathbb{P}(X_0 = s_i) \qquad i = 1, \ldots, n$$
2. P tale che $[P]_{ij} = p_{ij}$ è la *matrice di transizione* e possiede le seguenti proprietà:
   - P è una matrice quadrata di ordine n
   - $0 \leq p_{ij} \leq 1$
   - $\sum_{j=1}^{n} p_{ij} = 1 \qquad \forall\, i = 1, \ldots, n$

L'elemento p_{ij} di P è la probabilità subordinata di passare dallo stato i allo stato j. $(X_i)_{i\in\mathcal{N}}$ si può interpretare come un numero aleatorio che si evolve nel tempo, dato da $i \in \mathcal{N}$.

Si interpreta quindi l'elemento p_{ij} di P come la probabilità subordinata di passare dallo stato i allo stato j.

Si ha inoltre che

$$\begin{aligned}
\mathbb{P}(X_r = s_r | X_{r-1} = s_{r-1}, \ldots, X_0 = s_0) &= \frac{\mathbb{P}(X_r = s_r, X_{r-1} = s_{r-1}, \ldots, X_0 = s_0)}{\mathbb{P}(X_{r-1} = s_{r-1}, \ldots, X_0 = s_0)} \\
&= \frac{\rho_{s_0} p_{s_0,s_1} p_{s_1,s_2} \cdots p_{s_{r-1},s_r}}{\rho_{s_0} p_{s_0,s_1} p_{s_1,s_2} \cdots p_{s_{r-2},s_{r-1}}} \\
&= p_{s_{r-1},s_r}
\end{aligned}$$

La probabilità di trovarsi al tempo r (X_r) nello stato s_r sapendo tutti gli stati precedenti dipende solo dallo stato immediatamente precedente (*indipendenza dal passato*).

Esempio 5.1 (Passeggiata aleatoria) Una passeggiata aleatoria è rappresentata da una catena di Markov con matrice di transizione

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1-p & 0 & p & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1-p & 0 & p \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Lo stato iniziale e quello finale presentano condizioni al bordo assorbenti. Nel caso di $p = \frac{1}{2}$ si parla di *passeggiata aleatoria simmetrica.*

Esempio 5.2 Consideriamo due urne A e B, ciascuna con N palline. Fra tutte le palline ve ne sono N bianche ed N nere. Si scelgono una pallina dall'urna A e una pallina dall'urna B e si scambiano.

X_i rappresenta il numero di palline bianche nell'urna A al tempo i. L'insieme degli stati è quindi

$$S = \{0, 1, \dots, N\}$$

p_{kl} è la probabilità di passare dallo stato k allo stato l. Si ottiene

$$\begin{aligned} p_{k,k} &= \text{estraggo 2 palline bianche oppure 2 nere} \\ &= \frac{k}{N}\frac{N-k}{N} + \frac{N-k}{N}\frac{k}{N} \end{aligned}$$

$$\begin{aligned} p_{k,k+1} &= \text{estraggo 1 nera da } A \text{ e 1 bianca da } B \\ &= \frac{N-k}{N}\frac{N-k}{N} \\ &= \frac{(N-k)^2}{N^2} \end{aligned}$$

$$\begin{aligned} p_{k,k-1} &= \text{estraggo 1 bianca da } A \text{ e 1 nera da } B \\ &= \frac{k}{N}\frac{k}{N} \\ &= \frac{k^2}{N^2} \end{aligned}$$

Queste formule valgono anche per $k = 0$ e $k = N$. Si otiene quindi la matrice di transizione

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \frac{1}{N^2} & \frac{2(N-1)}{N^2} & \left(\frac{N-1}{N}\right)^2 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

5.1.1 La probabilità di transizione in n passi

Usando la formula delle probabilità composte possiamo calcolare la probabilità di passare dallo stato s allo stato s allo stato s' in n passi.

$$\begin{aligned}
\mathbb{P}(X_{m+n} = s'|X_m = s) &= \mathbb{P}(X_{m+n} = s'|X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0) \\
&= \frac{\mathbb{P}(X_{m+n} = s', X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0)}{\mathbb{P}(X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0)} \\
&= \frac{\sum_{s_{m+1},\cdots,s_{m+n-1}} \mathbb{P}(X_{m+n} = s', X_{m+n-1} = s_{m+n-1}, \ldots, X_0 = s_0)}{\mathbb{P}(X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0)} \\
&= \frac{\sum_{s_{m+1},\cdots,s_{m+n-1}} \rho_0\, p_{s_0,s_1} \cdots p_{s_{m-1},s}\, p_{s,s_{m+1}} \cdots p_{s_{m+n-1},s'}}{\rho_0\, p_{s_0,s_1}\, p_{s_1,s_2} \cdots p_{s_{m-1},s}} \\
&= \sum_{s_{m+1},\cdots,s_{m+n-1}} p_{s,s_{m+1}} \cdots p_{s_{m+n-1},s'} \\
&= [P^n]_{s,s'}
\end{aligned}$$

Tale probabilità *non* dipende da m, ma solo da n, ovvero dal numero di passi intermedi, e si ottiene come l'elemento di coordinate s, s' dell'n-esima potenza della matrice di transizione P; questo, da ora in avanti, sarà indicato nel modo seguente:

$$\mathbb{P}(X_{m+n} = s'|X_m = s) = [P^n]_{s,s'} \triangleq p^{(n)}_{s,s'}$$

Per convenzione, si definisce

$$p^{(0)}_{s,s'} = \rho_{s,s'} = \begin{cases} 1 & \text{se } s = s' \\ 0 & \text{altrimenti} \end{cases}$$

Per esempio

$$\mathbb{P}(X_{m+2} = s'|X_m = s) = \sum_{s_1} p_{s,s_1}\, p_{s_1,s'} = \left[P^2\right]_{s,s'}$$

5.1.2 Classi di equivalenza

Sia $(X_i)_{i\in\mathcal{N}}$ una catena di Markov con numero finito di stati ed omogenea. Uno stato s *comunica* con s' se

$$\exists\, n \,|\, p^{(n)}_{s,s'} > 0$$

ovvero se esiste un percorso di lunghezza tale che si passa con probabilità positiva dallo stato s allo stato s'. Questa proprietà si indica: $s \prec s'$.

Due stati, s e s', si dicono *equivalenti* se $s \prec s'$ e $s' \prec s$. Tale relazione è una relazione di equivalenza in quanto è riflessiva, simmetrica e transitiva. Basta verificare la proprietà transitiva:

$s \prec s' \Rightarrow \exists\, n_1 \mid p^{(n_1)}_{s,s'} > 0$

$s' \prec s'' \Rightarrow \exists\, n_2 \mid p^{(n_2)}_{s',s''} > 0$

Se si sceglie $n = n_1 + n_2$

$$p^{(n_1+n_2)}_{s,s''} = \left[P^{n_1+n_2}\right]_{s,s''} = \sum_{s_1} p^{(n_1)}_{s,s_1}\, p^{(n_2)}_{s_1,s''} \geq \underbrace{p^{(n_1)}_{s,s'}}_{>0}\, \underbrace{p^{(n_2)}_{s',s''}}_{>0} > 0$$

La relazione $\prec$ di comunicazione fra stati si può estendere senza ambiguità alle classi di equivalenza. Diciamo che $[s]$ comunica con $[s']$ e scriviamo $[s] \prec [s']$ se $s \prec s'$. È facile vedere che questa definizione non è ambigua, cioè non dipende dalla scelta degli elementi nelle classi di equivalenza.

Una classe di equivalenza si dice *massimale* se non è seguita da nessun'altra classe nella relazione che abbiamo definito. Se la catena di Markov si trova in uno stato di una classe massimale, agli istanti successivi, con probabilità 1, essa si trova in uno stato che appartiene alla stessa classe.

Si vuol definire il *periodo* di uno stato. Consideriamo:

$$A^+_s = \{\, n \,|p^{(n)}_{s,s} > 0\}$$

Il periodo di uno stato s è dato dal minimo comun divisore (MCD) degli elementi di A^+_s. Se il periodo è 1, si dice che lo stato è *aperiodico*. Per esempio, nella passeggiata aleatoria gli stati hanno periodo 2. Se l'insieme $A^+_s = \{0\}$, il periodo si può considerare ∞ o non definito, a seconda delle convenzioni.

Tutti gli stati di una classe di equivalenza hanno lo stesso periodo, per cui si può parlare del periodo di una classe di equivalenza.

DIMOSTRAZIONE. Sia $s \sim s'$, siano q periodo di s e q' periodo di s'. Basta dimostrare che q' divide ogni $n \in A^+_s$. Per l'equivalenza, si ha che:

$\exists\, n_1$ tale che $p^{(n_1)}_{s,s'} > 0$

$\exists\, n_2$ tale che $p^{(n_2)}_{s',s} > 0$

allora $(n_1 + n_2) \in A^+_s$ perché

$$\begin{aligned} p^{(n_1+n_2)}_{s,s} &= \sum_{s_1} p^{(n_1)}_{s,s_1}\, {}^{(n_2)}_{s_1,s} \\ &\geq p^{(n_1)}_{s,s'}\, p^{(n_2)}_{s',s} \\ &> 0 \end{aligned}$$

Ma $(n_1 + n_2)$ appartiene anche ad $A^+_{s'}$; quindi q e q' dividono $(n_1 + n_2)$. Inoltre, per ogni $n \in A^+_s$, $(n + n_1 + n_2)$ sta in $A^+_{s'}$ perché

$$p^{(n+n_1+n_2)}_{s',s'} \geq p^{(n_2)}_{s',s}\, p^{(n)}_{s,s}\, p^{(n_1)}_{s,s'} > 0$$

Quindi, q e q' dividono $(n+n_1+n_2)$ per ogni $n \in A_s^+$ e per ogni $n \in A_{s'}^+$, ovvero q e q' dividono n per ogni $n \in A_s^+$ e per ogni $n \in A_{s'}^+$. Questo implica che

$$q = q'$$

□

Una classe C di equivalenza di periodo $q < \infty$ si può decomporre in q sottoinsiemi:

$$C = C_0 \cup C_1 \cup \cdots \cup C_{q-1}$$

con la proprietà che se $s \in C_i$, $s' \in C_j$ e $p_{s,s'}^{(n)} > 0$ allora $n \equiv (j-i) \mod q$.

Se una classe di equivalenza massimale ha periodo q, i q sottoinsiemi in cui è decomposta sono percorsi ciclicamente dalla catena di Markov: cioè se $X_0 \in C_i$, allora $X_1 \in C_{[i+1]_q}$, $X_2 \in C_{[i+2]_q}$ dove indichiamo con la notazione $[k]_q$ l'elemento dell'insieme $\{0, \ldots, q-1\}$ equivalente a k modulo q.

5.2 Teorema ergodico

Si vuole studiare il comportamento della catena al passare del tempo.

Una catena di Markov con una sola classe di equivalenza aperiodica e con spazio degli stati finito ha la proprietà di convergenza ad una distribuzione invariante sugli stati e indipendente dallo stato iniziale. Questa proprietà è il risultato del seguente teorema detto *ergodico*.

Teorema 5.3 (Teorema ergodico) *Sia $(X_i)_{i \in \mathcal{N}}$ una catena di Markov omogenea con insieme finito di stati. Se la catena è irriducibile (ovvero ha una sola classe di equivalenza) aperiodica, allora esiste una distribuzione di probabilità $\Pi = \{\pi_1, \ldots, \pi_n\}$ sullo spazio degli stati, e delle costanti C e $0 \leq \delta < 1$ tali che $\forall s \in S$, $\forall n$ si ha:*

$$|p_{s,s'}^{(n)} - \pi_{s'}| \leq C\delta^n$$

In altre parole esistono π_s tali che:

1. $0 \leq \pi_s \leq 1$
2. $\sum_{s \in S} \pi_s = 1$

e vale che

$$\lim_{n \to +\infty} p_{s's}^{(n)} = \pi_s \qquad \forall s' \in S$$

con velocità esponenziale.

Questo teorema può essere utilizzato anche nel caso di periodo q maggiore di 1, considerando la catena di Markov associata alla matrice di transizione P^q. Infatti, la restrizione di tale catena ad ognuno dei sottoinsiemi $C_0, C_1, \ldots, C_{q-1}$ soddisfa le ipotesi del teorema ergodico.

La distribuzione di probabilità Π che appare nel teorema ergodico ha la proprietà di essere una distribuzione *invariante*: se la poniamo come distribuzione iniziale, se cioè $\mathbb{P}(X_0 = s) = \pi_s$ valesse per ogni $s \in S$, allora per ogni $s \in S$ e per ogni $n \geq 0$

$$\mathbb{P}(X_n = s) = \pi_s$$

Questa proprietà ci permette di calcolare π_s come soluzione di un sistema di equazioni lineari. Infatti

$$\begin{aligned} \pi_s &= \mathbb{P}(X_1 = s) \\ &= \sum_{s' \in S} \mathbb{P}(X_0 = s')\, p_{s',s} \\ &= \sum_{s' \in S} \pi_{s'}\, p_{s',s} \end{aligned}$$

Inoltre, dato che π_s è una distribuzione di probabilità, abbiamo

$$\sum_{s \in S} \pi_s = 1$$

(π_s) si dice *distribuzione stazionaria* per la catena. Si può dimostrare che, sotto le ipotesi del teorema ergodico, esiste una e una sola soluzione di questo sistema di $|S| + 1$ equazioni in $|S|$ incognite; una delle equazioni, in questo caso una delle prime $|S|$ equazioni, è funzione lineare delle altre e quindi può non essere considerata nella soluzione del sistema:

$$(1) \begin{cases} \Pi = \Pi^t P \\ \sum \pi_s = 1 \end{cases}$$

dove si è posto

$$\Pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_n \end{pmatrix}$$

Il teorema ergodico ci dice che la catena dimentica lo stato di partenza all'aumentare del tempo. Si dimostra ora l'*unicità*.

DIMOSTRAZIONE. Supponiamo che $(\mu_s)_{s \in S}$ sia un'altra distribuzione che soddisfa il sistema (1). Si ottiene

$$\begin{cases} \mu = \mu^t P \\ \sum \mu_s = 1 \end{cases}$$

si ha che

$$\mu = \mu^t P \quad \Rightarrow \quad \mu = \mu^t P = \mu^t P^2 = \cdots = \mu^t P^n$$

se n cresce all'infinito, P^n converge alla matrice

$$\begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$$

ovvero

$$\mu_j = \sum_i \mu_i p_{ij} = \sum_i \mu_i p_{ij}^{(n)}$$

Prendendo il limite $\lim_{n \to +\infty} p_{ij}^{(n)} = \pi_j$, si ottiene

$$\mu_j = \sum_i \mu_i \pi_j = \pi_j \underbrace{\sum_i \mu_i}_{1}$$

Ne segue che

$$\mu_j = \pi_j$$

□

Capitolo 6

Catene di Markov a tempo continuo

Consideriamo il caso di un ufficio postale con n sportelli. In un tempo t arriveranno dei clienti e degli altri verranno serviti.

Si vuole studiare qual è la distribuzione di probabilità che nel sistema siano presenti n clienti. Chiaramente, per fare ciò non basta considerare un solo numero aleatorio perchè il numero di clienti si considera a vari istanti nel tempo. Si introduce quindi un nuovo concetto, quello di processo stocastico.

Un *processo stocastico* $(X_t)_{t \in I}$, con I intervallo di $\mathbb{R}$, è una famiglia di numeri aleatori indicizzati su un intervallo I di $\mathbb{R}$.

Se si ritorna all'esempio iniziale, il numero di clienti nel sistema al tempo t sarà dato da un numero aleatorio X_t e, al variare del tempo, dal processo stocastico $(X_t)_{t \geq 0}$.
Una catena di Markov $(X_i)_{i \in \mathcal{N}}$ è un esempio di processo stocastico a *tempi discreti* in quanto gli indici variano fra i numeri naturali.

6.1 Catene di Markov a tempo continuo con insieme di stati numerabile

Una catena di Markov omogenea in tempo continuo con insieme di stati numerabile è un processo stocastico $(X_t)_{t \geq 0}$ tale che $I(X_t) = \mathcal{N}$ per ogni $t \geq 0$. Nel caso di catene di Markov omogenee in tempo continuo, bisogna considerare per ogni intervallo di tempo t una matrice di transizione $p_{s,s'}(t) = [\Pi(t)]_{ss'}$ non essendovi un intervallo di tempo minimale come nel caso del tempo discreto. Le matrici di transizione sono collegate fra loro dalle equazioni di Chapman-Kolmogorov che si possono scrivere in maniera sintetica

$$\Pi(t+h) \;=\; \Pi(t)\,\Pi(h) \qquad \forall\, t, h \geq 0$$

o esplicitamente

$$p_{s,s'}(t+t') = \sum_{s''} p_{s,s''}(t)\, p_{s'',s'}(t')$$

Per poter trattare esempi interessanti, per esempio la teoria delle code, dobbiamo considerare il caso in cui lo spazio degli stati sia al più numerabile. In questo caso $\Pi(t)$ è una "matrice" con infinite righe e infinite colonne con elementi non negativi tale che la somma delle serie degli elementi di ogni riga sia uguale a 1.

Il prodotto riga per colonna di due matrici di questo tipo si può definire come nel caso di matrici quadrate con una serie al posto della somma: il risultato è ancora una matrice con la stessa proprietà, come è facile verificare.

Come nel caso del tempo discreto si possono determinare le probabilità di transizione in più passi da quelle in un passo, così nel caso di tempo continuo è possibile determinare le probabilità di transizione a partire dal loro comportamento per un intervallo di tempo infinitamente piccolo. Il caso più semplice si ha nel modello denominato *"processo di Poisson"*.

6.2 Processi di Poisson

Il processo di Poisson è una catena di Markov a tempo continuo con spazio degli stati $S = \mathcal{N}$ e probabilità di transizione $p_{s,s'}(t)$ che verificano

1. $p_{s,s}(h) \;=\; 1 - \lambda h + o(h)$
2. $p_{s,s+1}(h) \;=\; \lambda h + o(h)$
3. $p_{s,s'}(h) \;=\; o(h) \qquad \text{per } s' \notin \{s, s+1\}$

dove λ è un parametro strettamente positivo e $o(h)$ denota infinitesimi di ordine maggiore di h, che supponiamo essere uniformemente infinitesimi.

A partire da queste ipotesi possiamo ottenere un sistema di infinite equazioni differenziali dette *equazioni di Kolmogorov in avanti* per le probabilità di transizione nei vari stati a partire da uno stato fissato $(\bar{s})$. Fissiamo per esempio $\bar{s} = 0$. Intuitivamente, questo significa che al tempo 0 non c'è nessun cliente nel sistema. Si pone

$$\mu_s(t) \;=\; p_{0,s}(t) \qquad \text{per } s \in S$$

Le μ_s verificano il sistema di equazioni:

$$\begin{cases} \mu_0'(t) \;=\; -\lambda \mu_0(t) \\ \mu_s'(t) \;=\; -\lambda \mu_s(t) + \lambda \mu_{s-1}(t) \qquad \text{per } s \geq 1 \end{cases}$$

Infatti calcoliamo il rapporto incrementale $\dfrac{\mu_s(t+h) - \mu_s(t)}{h}$ per $s > 0$.

$$\frac{\mu_s(t+h) - \mu_s(t)}{h} =$$

$$\frac{p_{0,s}(t+h) - p_{0,s}(t)}{h} =$$

$$\frac{\sum_{j\in S} p_{0,j}(t)p_{j,s}(h) - p_{0,s}(t)}{h} =$$

$$\frac{(1-\lambda h)p_{0,s}(t) + \lambda h p_{0,s-1}(t) + \sum_{j\in S, j\neq s, j\neq s-1} p_{0,j}(t)p_{j,s}(h) - p_{0,s}(t)}{h} =$$

$$-\lambda p_{0,s}(t) + \lambda p_{0,s-1}(t) + \frac{o(h)}{h} =$$

$$-\lambda \mu_s(t) + \lambda \mu_{s-1}(t) + \frac{o(h)}{h}$$

Per h tendente a zero, tale rapporto converge a

$$\mu_s'(t) \;=\; -\lambda \mu_s(t) + \lambda \mu_{s-1}(t) \quad \text{per } s \geq 1$$

Per $s = 0$, si ottiene

$$\frac{\mu_0(t+h) - \mu_0(t)}{h} =$$

$$\frac{p_{0,0}(t+h) - p_{0,0}(t)}{h} =$$

$$\frac{\sum_{j\in S} p_{0,j}(t)p_{j,0}(h) - p_{0,0}(t)}{h} =$$

$$\frac{(1-\lambda h)p_{0,0}(t) + \sum_{j\in S, j\neq 0} p_{0,j}(t)p_{j,0}(h) - p_{0,0}(t)}{h} =$$

$$-\lambda p_{0,0}(t) + \frac{o(h)}{h} =$$

$$-\lambda \mu_0(t) + \frac{o(h)}{h}$$

che converge a

$$\mu_0'(t) \;=\; -\lambda \mu_0(t)$$

per h che tende a zero. Possiamo risolvere iterativamente il sistema di equazioni ottenendo

$$p_{0,s}(t) \;=\; \frac{(\lambda t)^s}{s!}\, e^{-\lambda t}$$

La distribuzione del processo a tempo t è di Poisson con parametro λt.

Nel caso di un $\bar{s}$ arbitrario si ha una traslazione di $\bar{s}$:

$$\begin{cases} p_{\bar{s},s}(t) = 0 & \text{per } s < \bar{s} \\ p_{\bar{s},s}(t) = \dfrac{(\lambda t)^{s-\bar{s}}}{(s-\bar{s})!} e^{-\lambda t} & \text{per } s \geq \bar{s} \end{cases}$$

Si vede dalle probabilità di transizione che il processo di Poisson è non decrescente con probabilità 1. Possiamo rappresentare graficamente il processo di Poisson nel seguente modo
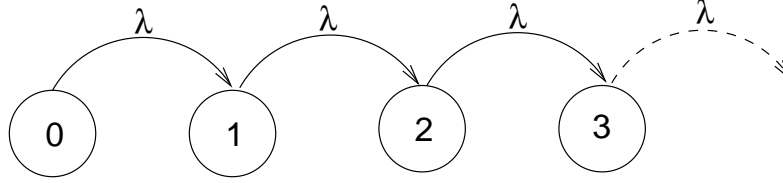
Figura 6.1: Schema di un generico processo di Poisson con stato iniziale in 0.

dove la freccia che collega due stati sovrascritta da λ indica che l'intensità di transizione fra i due stati è λ. Osserviamo che da ogni punto s parte una freccia e che in ogni punto $s' \geq 1$ arriva una freccia. Queste due frecce, una uscente ed una entrante, corrispondono ai due termini nella parte destra dell'equazione differenziale. Per $s = 0$ si ha solo una freccia uscente che corrisponde all'unico termine dell'equazione differenziale.

Se indichiamo con $P_s(t) = \mathbb{P}(N(t) = s)$ la probabilità che il processo di Poisson si trovi nello stato s al tempo t, abbiamo

$$P_s(t) = \sum_{\bar{s} \in S} \rho_{\bar{s}} \, p_{\bar{s},s}(t)$$

dove ρ_s è la distribuzione iniziale. Da qui si vede che per ogni distribuzione iniziale le $P_s(t)$ soddisfano lo stesso sistema di equazioni

$$\begin{cases} P_0'(t) = -\lambda P_0(t) \\ P_s'(t) = -\lambda P_s(t) + \lambda P_{s-1}(t) & \text{per } s \geq 1 \end{cases}$$

Le $p_{\bar{s},s}(t)$ si possono considerare casi particolari in cui $\rho_{\bar{s}} = 1$ e $\rho_s = 0$ per $s \neq \bar{s}$.

6.3 Processi a coda

Consideriamo ora alcuni esempi di catene di Markov in tempo continuo che servono come modelli per i processi a coda. Nella teoria delle code vi è una notazione simbolica per denotare i tipi di processi. Gli esempi che considereremo descrivono un flusso di arrivi che segue un processo di Poisson con parametro λ. I clienti che trovano uno sportello libero iniziano un tempo di servizio allo sportello, gli altri si mettono in coda. Quando uno sportello si libera, uno dei clienti in coda inizia il suo servizio.

Per quello che ci interessa non ha importanza in quale ordine i clienti accedono al servizio; poniamo, per esempio, che l'ordine sia casuale. Facciamo l'ipotesi che i tempi di servizio siano indipendenti e identicamente distribuiti e indipendenti dal processo di Poisson che regola il flusso degli arrivi. Supponiamo che la distribuzione del tempo di servizio sia esponenziale di parametro μ.

Un processo di questo tipo viene denotato col simbolo $M/M/n$. Il primo M indica che il flusso degli arrivi è di Poisson, il secondo indica che il tempo di servizio è esponenziale ed n indica il numero degli sportelli e può variare da 1 a ∞ (anche il valore ∞ è ammissibile).

6.3.1 Code $M/M/\infty$

Si considera una situazione idealizzata in cui vi sono infiniti sportelli. Il processo degli arrivi è dato da un processo di Poisson di parametro λ ed il tempo di servizio T ha distribuzione esponenziale di parametro μ.

Sia $(X_t)_{t\geq 0}$ il processo che indica il numero di persone presenti nel sistema. Come distribuzione iniziale si assume che

$$\begin{cases} \mathbb{P}(X_0 = 0) = 1 & \\ \mathbb{P}(X_0 = i) = 0 & i > 0 \end{cases}$$

Inoltre, il tempo di servizio ed il processo degli arrivi sono stocasticamente indipendenti. Per conoscere l'intensità di servizio si calcola la probabilità che un cliente venga servito nel tempo $(t+h)$ se non è stato servito fino al tempo t.

$$\begin{aligned} \mathbb{P}(T < t+h | T > t) &= \frac{\mathbb{P}(t < T < t+h)}{\mathbb{P}(T > t)} \\ &= \frac{e^{-\mu t} - e^{-\mu(t+h)}}{e^{-\mu t}} \\ &= 1 - e^{-\mu h} \\ &= 1 - (1 - \mu h + o(h)) \\ &= \mu h + o(h) \end{aligned}$$

Supponiamo che nel sistema vi siano n clienti. Se nessuno di essi è stato servito fino al tempo t, la probabilità che almeno uno di essi sia servito nel tempo $(t+h)$ per h piccolo è allora

$$\begin{aligned} 1 - \mathbb{P}(T_1 > t+h, \ldots, T_n > t+h | T_1 > t, \ldots, T_n > t) &= \\ 1 - \mathbb{P}(T > t+h | T > t)^n &= \\ 1 - e^{-n\mu h} &= \\ n\mu h + o(h) & \end{aligned}$$

L'intensità con cui si esce dal sistema è quindi proporzionale al numero di clienti che si trovano presenti nel sistema stesso, ovvero allo stato in cui si trova la catena. Si ottiene che il processo può essere rappresentato graficamente nel seguente modo:
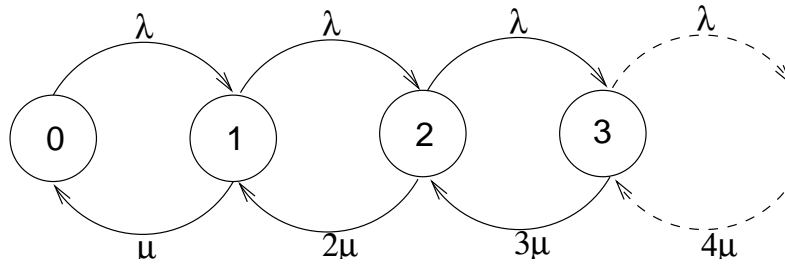
Figura 6.2: Rappresentazione grafica di una coda $M/M/\infty$.

Ponendo $p_{0,s}(t) = \mu_s(t)$, si possono scrivere le equazioni di Kolmogorov utilizzando la regola descritta precedentemente:

$$\begin{cases} \mu_0'(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \mu_i'(t) = -(\lambda + i\mu)\,\mu_i(t) + \lambda\,\mu_{i-1}(t) + (i+1)\mu\,\mu_{i+1}(t) \end{cases}$$

Si ricerca una soluzione stazionaria $(p_i)_{i\geq 0}$ di tale sistema, ovvero una distribuzione che *non* dipende dal tempo. Si impone che

$$p_i' = 0$$

ottenendo

$$\begin{cases} 0 = \mu p_1 - \lambda p_0 \\ 0 = -(\lambda + i\mu)p_i + \lambda p_{i-1} + (i+1)\mu p_{i+1} \\ \sum_{i=0}^{+\infty} p_i = 1 \end{cases}$$

Sommando le equazioni fino alla i-esima si ottiene la formula ricorsiva

$$p_i = \frac{\lambda}{i\mu} p_{i-1} = \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i p_0$$

Deve essere $\sum p_i = 1$, quindi

$$\sum \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i p_0 = 1$$

La serie $\sum \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i$ è la serie esponenziale con argomento $e^{\frac{\lambda}{\mu}}$, quindi

$$p_i = \frac{\lambda}{i\mu} p_{i-1} = \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i e^{-\frac{\lambda}{\mu}}$$

In questo caso la distribuzione stazionaria esiste sempre.

6.3.2 Code $M/M/1$

Anche in questo caso, si suppone che il tempo di servizio T abbia distribuzione esponenziale di parametro μ, che sia indipendente dal processo degli arrivi e che questi siano regolati da un processo di Poisson di parametro λ.
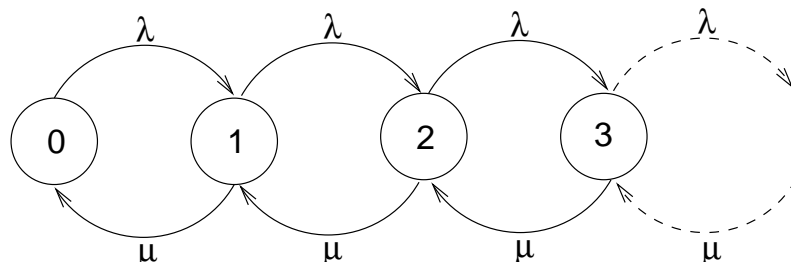
Figura 6.3: Schema di una coda $M/M/1$ con stato iniziale in 0.

Poiché si ha un solo sportello, l'intensità con cui il numero di clienti diminuisce è costante e pari a μ. Il grafico del processo è

ed il sistema di equazioni differenziali, posto $\mu_s = p_{\bar{s},s}(t)$, è dato da

$$\begin{cases} \mu_0'(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \\ \mu_i'(t) = -(\lambda+\mu)\,\mu_i(t) + \lambda\,\mu_{i-1}(t) + \mu\,\mu_{i+1}(t) \end{cases}$$

Anche in questo caso si cerca la *soluzione stazionaria*.

$$\begin{cases} 0 = \mu p_1 - \lambda p_0 \\ \\ 0 = -(\lambda+\mu)p_i + \lambda p_{i-1} + \mu p_{i+1} \\ \\ \sum_{i=0}^{+\infty} p_i = 1 \end{cases}$$

Dal sistema precedente si ottiene la relazione ricorsiva

$$p_i = \frac{\lambda}{\mu}p_{i-1} = \left(\frac{\lambda}{\mu}\right)^i p_0$$

Quindi

$$\sum_{i=0}^{+\infty} p_i = 1 \Rightarrow \underbrace{\left(\sum_{i\geq 0}\left(\frac{\lambda}{\mu}\right)^i\right)}_{\substack{\text{distribuzione}\\ \text{geometrica}\\ \text{parametro } \lambda/\mu}} p_0 = 1$$

La soluzione dell'equazione è

$$p_0 = \left(1 - \frac{\lambda}{\mu}\right)$$

per cui la distribuzione di probabilità diventa

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right)$$

Ovvero, p_i è una distribuzione geometrica di parametro $\dfrac{\lambda}{\mu}$ *se*

$$\sum \left(\frac{\lambda}{\mu}\right)^i < +\infty$$

cioè se $\frac{\lambda}{\mu} < 1$, quindi $\lambda < \mu$, ovvero l'intensità di arrivo deve essere più bassa di quella di servizio.

Capitolo 7

Statistica

7.1 Densità subordinata di due numeri aleatori

Prima di affrontare lo studio della statistica, si introduce la densità subordinata di un numero aleatorio Y rispetto ad un numero aleatorio X. Sia $f(x, y)$ la densità congiunta di (X, Y). Si dice *densità subordinata di Y rispetto ad X*:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

se $f_X(x) > 0$.

Si ottiene subito una formula di Bayes per la densità. Poiché vale

$$f(x,y) = f_{Y|X}(y|x)\, f_X(x)$$

ma anche

$$f(x,y) = f_{X|Y}(x|y)\, f_Y(y)$$

si ottiene

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)\, f_Y(y)}{f_X(x)}$$

7.2 Statistica Bayesiana

7.2.1 Induzione statistica sulla Binomiale

Si consideri una successione di eventi $(E_i)_{i\in\mathcal{N}}$ stocasticamente indipendenti subordinatamente alla conoscenza di un parametro Θ, ovvero tali che

$$\mathbb{P}(E_i = 1|\Theta = \theta) = \theta$$

dove $0 < \theta < 1$.

Gli eventi E_i possono essere pensati come il risultato di un esperimento; il fatto che essi siano stocasticamente indipendenti rispetto alla conoscenza di Θ significa che

$$\mathbb{P}(E_1 = \epsilon_1, \dots, E_n = \epsilon_n|\Theta = \theta) = \prod_{i=1}^{n} \mathbb{P}(E_i = \epsilon_i|\Theta = \theta)$$

dove $\epsilon_i \in \{0,1\}$.

Si assume che Θ abbia una *distribuzione a priori* $\pi_0(\theta)$ e si vuole vedere come cambia la distribuzione di Θ dopo aver effettuato n esperimenti $E_1, \ldots, E_n$. Supponiamo quindi di aver effettuato gli n esperimenti con esito $E_1 = \epsilon_1, \ldots, E_n = \epsilon_n$. La *densità a posteriori* di Θ è data da

$$\pi_n(\theta | E_1 = \epsilon_1, \ldots, E_n = \epsilon_n)$$

Usando la formula di Bayes per la densità subordinata di due numeri aleatori si ottiene

$$\begin{aligned}
\pi_n(\theta | E_1 = \epsilon_1, \ldots, E_n = \epsilon_n) &= \frac{\mathbb{P}(E_1 = \epsilon_1, \ldots, E_n = \epsilon_n | \Theta = \theta)\, \pi_0(\theta)}{\mathbb{P}(E_1 = \epsilon_1, \ldots, E_n = \epsilon_n)} \\
&= \frac{1}{c}\, \pi_0(\theta) \prod_{i=1}^{n} \mathbb{P}(E_i = \epsilon_i | \Theta = \theta)
\end{aligned}$$

dove $c = P(E_1 = \epsilon_1, \ldots, E_n = \epsilon_n)$.

In particolare, se la distribuzione a priori $\pi_0(\theta)$ è una distribuzione beta $\mathcal{B}(\alpha, \beta)$ di parametri α e β, anche la distribuzione a posteriori sarà dello stesso tipo , di parametri

$$\alpha' = \alpha + \textstyle\sum_{i=1}^{n} \epsilon_i \qquad\qquad \beta' = \beta + n - \textstyle\sum_{i=1}^{n} \epsilon_i$$

dove $\sum_{i=1}^{n} \epsilon_i$ e $n - \sum_{i=1}^{n} \epsilon_i$ contano, rispettivamente, il numero di eventi verificati e il numero degli eventi che non si sono verificati. Infatti

$$\begin{aligned}
\pi_n(\theta | E_1 = \epsilon_1, \ldots, E_n = \epsilon_n) &= \frac{1}{c}\, \pi_0(\theta) \prod_{i=1}^{n} \mathbb{P}(E_i = \epsilon_i | \Theta = \theta) \\
&\sim \theta^{\alpha-1} (1-\theta)^{\beta-1}\, \theta^{\sum_i \epsilon_i} (1-\theta)^{n - \sum_i \epsilon_i} \\
&= \theta^{\alpha - 1 + \sum_i \epsilon_i} (1-\theta)^{\beta - 1 + n - \sum_i \epsilon_i}
\end{aligned}$$

dove il segno $\sim$ indica che la densità è proporzionale alla funzione di Θ sopraindicata.

Si ricorda la densità della distribuzione beta

$$\pi_0(\theta) = \begin{cases} K\, \theta^{\alpha-1} (1-\theta)^{\beta-1} & \theta \in [0,1] \\ 0 & \textit{altrimenti} \end{cases}$$

dove K si calcola come

$$K = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}$$

Ne segue che la densità a posteriori è una distribuzione beta di parametri $\alpha' = \alpha - 1 + \sum\limits_i \epsilon_i$ e $\beta' = \beta - 1 + n - \sum\limits_i \epsilon_i$, ovvero

$$\pi_n(\theta | E_1 = \epsilon_1, \ldots, E_n = \epsilon_n) = \begin{cases} \frac{\Gamma(\alpha'+\beta')}{\Gamma(\alpha')\,\Gamma(\beta')}\, \theta^{\alpha'-1} (1-\theta)^{\beta'-1} & \theta \in [0,1] \\ 0 & \textit{altrimenti} \end{cases}$$

7.2.2 Induzione statistica sulla media della distribuzione normale

Sia $(X_i)_{i\in\mathcal{N}}$ una successione di numeri aleatori stocasticamente indipendenti subordinatamente alla conoscenza di un parametro Θ con densità subordinata

$$f(x_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x_i-\theta)^2}{2\sigma^2}}$$

con $\sigma \in \mathbb{R}^+$.

Se la distribuzione a priori $\pi_0(\theta)$ è una gaussiana $N(\mu_0, \sigma_0^2)$, dopo aver ottenuto i risultati dei primi n esperimenti la distribuzione a posteriori è allora

$$\begin{aligned}
\pi_n(\theta|X_1,\dots,X_n) &= \pi_0(\theta)\, f(x_1,\dots,x_n|\theta) \\
&= \pi_0(\theta) \prod_{i=1}^{n} f(x_i|\theta) \\
&\sim e^{-\dfrac{(\sigma-\mu_0)^2}{2\sigma_0^2}} e^{-\sum_{i=1}^{n} \dfrac{(x_i-\theta)^2}{2\sigma^2}} \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)\theta^2 - 2\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum x_i}{\sigma^2}\right)\theta\right] \\
&= \exp\left[-\frac{1}{2}\frac{(\theta-m_n)^2}{\sigma_n^2}\right]
\end{aligned}$$

dove $m_n = \dfrac{\dfrac{\mu_0}{\sigma_0^2}+\dfrac{\sum x_i}{\sigma^2}}{\dfrac{1}{\sigma_0^2}+\dfrac{n}{\sigma^2}}$ e $\sigma_n^2 = \left(\dfrac{1}{\sigma_0^2}+\dfrac{n}{\sigma^2}\right)^{-1}$. La densità a posteriori è quindi una gaussiana di parametri

$$N\left(\frac{\frac{\mu_0}{\sigma_0^2}+\dfrac{\sum x_i}{\sigma^2}}{\dfrac{1}{\sigma_0^2}+\dfrac{n}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2}+\dfrac{n}{\sigma^2}}\right)$$

Appendice A

Richiami di Analisi

In questa sezione si richiamano definizioni e concetti dell'analisi in una variabile fondamentali per lo sviluppo della teoria delle probabilità.

A.1 Limiti

A.1.1 Limite di una successione

Sia $(a_n)_{n \in \mathcal{N}}$ una successione di numeri reali. Essa si dice

1. *convergente* se
$$\lim_{n \to \infty} a_n = L < \infty$$
ovvero se $\forall\, \epsilon > 0 \;\; \exists\, N = N(\epsilon) \;\mid\; \forall n > N$
$$|a_n - L| < \epsilon$$

2. *divergente* se
$$\lim_{n \to \infty} a_n = +\infty$$
ovvero se
$\forall\, M > 0 \;\; \exists\, N = N(M) \;\mid\; \forall n > N$
$$a_n > M$$

Una successione può anche non essere né convergente né divergente. Per esempio, la successione $a_n = (-1)^n$ oscilla fra 1 e -1.

A.1.2 Limiti per le funzioni

Una funzione $f : \mathbb{R} \longrightarrow \mathbb{R}$ ha

1. *limite finito* in x
$$\lim_{y \to x} f(y) = L < \infty$$

Nell'analisi, tale scrittura ha il seguente significato:

$\forall \epsilon > 0 \;\; \exists \delta = \delta(\epsilon) \;\; | \;\; \forall |y - x| < \delta$

$$|f(y) - L| < \epsilon$$

2. *limite infinito* in x

$$\lim_{y \to x} f(y) \; = \; +\infty$$

Nell'analisi, tale scrittura ha il seguente significato:

$\forall M > 0 \;\; \exists \delta = \delta(M) \;\; | \;\; \forall |y - x| < \delta$

$$f(y) > M$$

A.1.3 Continuità

Una funzione si dice *continua* in un punto x_0 se

$$\lim_{x \to x_0^-} f(x) \; = \; \lim_{x \to x_0^+} f(x) = f(x_0)$$

dove $\lim_{x \to x_0^-} f(x)$, $\lim_{x \to x_0^+} f(x)$ si dicono rispettivamente *limite sinistro* e *limite destro* di f in quanto il primo di essi viene fatto per gli $x < x_0$, il secondo per gli $x > x_0$.

A.1.4 Limiti notevoli

Si ricordano i seguenti limiti notevoli:

1.
$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \; = \; e$$

2. $\forall x \in \mathbb{R}$
$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n \; = \; e^x$$

3.
$$\lim_{x \to 0} \frac{\log(1 + x)}{x} \; = \; 1$$

A.1.5 Serie notevoli

Si ricordano le seguenti serie notevoli:

1. La serie *geometrica*
$$\sum_{n=0}^{\infty} x^n \; = \; \frac{1}{1 - x}$$

per ogni $|x| < 1$.

2. La serie "derivata" della serie geometrica

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}$$

per ogni $|x| < 1$.

3. La serie *esponenziale*

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

per ogni $x \in \mathbb{R}$.

A.2 Le derivate

Funzione $f(x)$	**Derivata** $f'(x)$
x^n	$n\, x^{n-1}$
e^x	e^x
$\log\, x$	$\frac{1}{x}$
$\sin\, x$	$\cos\, x$
$\cos\, x$	$-\sin\, x$
$e^{-\frac{x^2}{2}}$	$-x\, e^{-\frac{x^2}{2}}$

A.2.1 Tabella delle principali regole di derivazione

$\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)+g(x)\right]$	$f'(x)+g'(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)\,g(x)\right]$	$f'(x)\,g(x)+f(x)\,g'(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{f(x)}{g(x)}\right]$	$\frac{f'(x)\,g(x)-f(x)\,g'(x)}{g^2(x)}$
$\frac{\mathrm{d}}{\mathrm{d}x}\left[f(g(x))\right]$	$f'(g(x))\cdot g'(x)$

A.3 Gli integrali

1. **Formula di integrazione per parti**

$$\int_a^b f(x)\,g'(x)\,\mathrm{d}x \;=\; \left[f(x)\,g(x)\right]_a^b - \int_a^b f'(x)\,g(x)\,\mathrm{d}x$$

2. **Cambio di variabili**

$$x = g(y) \Rightarrow dx = g'(y)\,\mathrm{d}y$$

$$\int_a^b f(x)\,\mathrm{d}x \;=\; \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))\,g'(y)\,\mathrm{d}y$$

Appendice B

Integrali bidimensionali

In questa appendice si richiamano in breve alcune nozioni di analisi delle funzioni multidimensionali necessarie per lo studio della teoria delle probabilità.

B.1 Area delle figure bidimensionali

Sia A una regione del piano; la sua area è data da

$$\text{area}\, A = \int\int_A \mathrm{d}x\mathrm{d}y$$
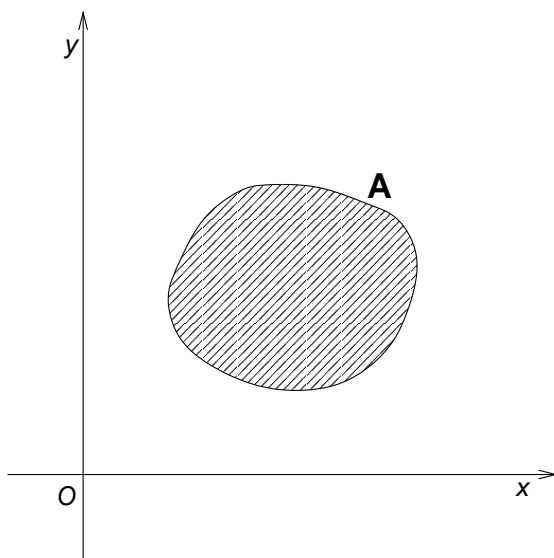
Figura B.1: Una generica regione del piano, indicata con A.

Analogamente al caso unidimensionale, in cui la lunghezza del segmento $[a, b]$ è data da

$$l([a,b]) = \int_a^b \mathrm{d}x$$

B.2 Integrale delle funzioni in due variabili

Sia $f : \mathbb{R}^2 \to \mathbb{R}$ ovvero $z = f(x, y)$. Una funzione in due variabili descrive una superficie in $\mathbb{R}^3$ di coordinate $(x,\, y,\, f(x, y))$. Si vuole calcolare il volume compreso fra la superficie descritta dalla funzione e il piano xy. Tale volume è dato dall'integrale

$$\int\int_{\mathbb{R}^2} f(x, y)\,\mathrm{d}x\mathrm{d}y$$

Si può anche considerare l'integrale su una regione $\Omega \subset \mathbb{R}^2$; intuitivamente, l'idea è che il volume del solido di base Ω e descritto dalla funzione $f(x, y)$ sia decomponibile in volumi infinitesimi tali che:

$$\text{volume}\,\Omega = \sum f(x, y)\,\Delta x \Delta y$$
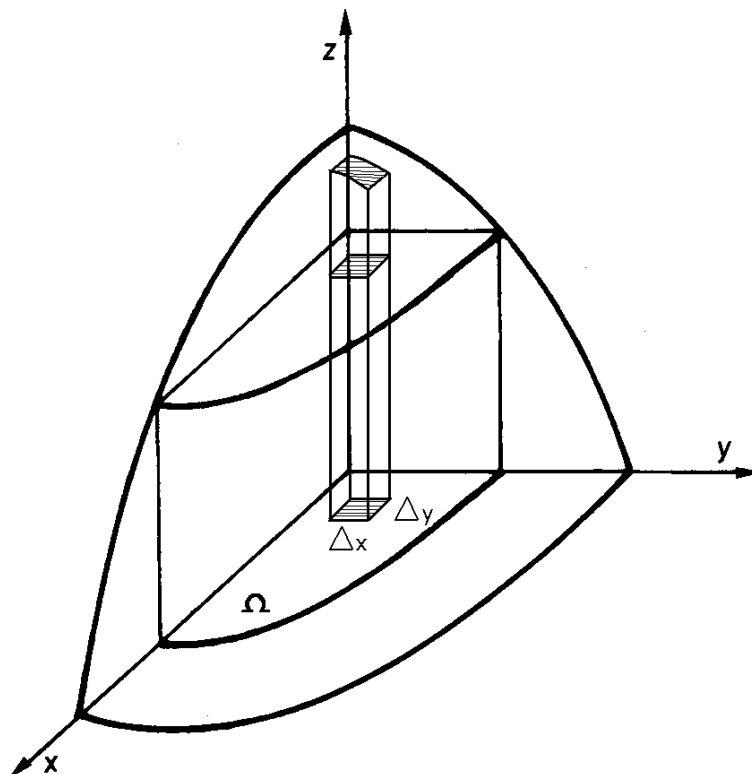
Figura B.2: Una funzione in due variabili integrata sull'area Ω del piano xy.

L'integrale è il limite per Δx e Δy che tendono a 0 di tali somme.

$$\sum \inf_{\Delta x \Delta y} f\,\Delta x \Delta y \,\leq\, I(f)_\Omega \,\leq\, \sum \sup_{\Delta x \Delta y} f\,\Delta x \Delta y$$

In pratica, per calcolare gli integrali doppi si usa il *teorema di Fubini-Tonelli* che ci permette di calcolarli come due integrali in una dimensione annidati l'uno dentro l'altro.

Esempio B.1 Sia $A = \{1 < x < 2,\, 3 < y < 4\}$

$$
\begin{aligned}
\int\int_A x^2 y \,\mathrm{d}x\mathrm{d}y &= \int_1^2 \mathrm{d}x \underbrace{\int_3^4 x^2 y \,\mathrm{d}y}_{x \text{ è un parametro!}} \\
&= \int_1^2 x^2 \left(\int_3^4 y \,\mathrm{d}y\right) \mathrm{d}x \\
&= \int_1^2 x^2 \left[\frac{1}{2}y^2\right]_3^4 \mathrm{d}x \\
&= \frac{7}{2}\int_1^2 x^2 \,\mathrm{d}x \\
&= \frac{49}{6}
\end{aligned}
$$

Esempio B.2 Sia $B = \{0 < x < 1,\, x-1 < y < x+1\}$
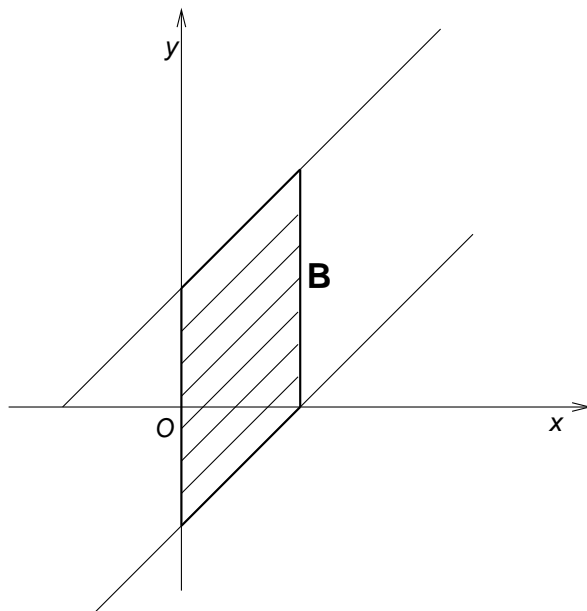
Figura B.3: La regione del piano individuata da B.

$$
\begin{aligned}
\int\int_B e^{-y} \,\mathrm{d}x\mathrm{d}y &= \int_0^1 \mathrm{d}x \int_{x-1}^{x+1} e^{-y} \,\mathrm{d}y \\
&= \int_0^1 \left[e^{-y}\right]_{x-1}^{x+1} \mathrm{d}x \\
&= \int_0^1 \left(e^{-(x+1)} - e^{-(x-1)}\right) \mathrm{d}x
\end{aligned}
$$
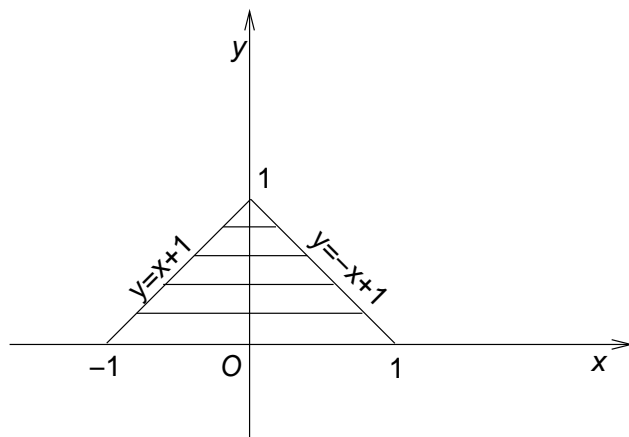
Figura B.4: Dominio individuato da D

Esempio B.3 A volte il problema è quello di suddividere opportunamente il dominio.

Cambiare l'ordine di integrazione, ovvero integrare in un ordine diverso da quello con cui sono specificati i differenziali delle variabili, non cambia il risultato dell'integrale ed a volte rende più semplice il calcolo dell'integrale, come nell'esempio della figura B.4, dove $D = \{0 < y < 1,\, y-1 < x < -y+1\}$.

$$\begin{aligned}\int\int_D f(x,y)\,\mathrm{d}x\mathrm{d}y &= \int_0^1 \mathrm{d}y \int_{y-1}^{-y+1} f(x,y)\,\mathrm{d}x \\ &= \int_0^1 \mathrm{d}x \int_0^{-x+1} f(x,y)\,\mathrm{d}y + \int_{-1}^0 \mathrm{d}x \int_0^{x+1} f(x,y)\,\mathrm{d}y\end{aligned}$$

Al primo passaggio, gli estremi di integrazione si trovano tracciando le parallele all'asse x e trovando i punti di intersezione di queste con il bordo della regione D. Al secondo passaggio invece, l'integrale è stato spezzato in due parti e gli estremi sono stati trovati con lo stesso metodo di prima tracciando però le parallele rispetto all'asse delle y.

B.3 Cambio di variabili

B.3.1 Derivate parziali rispetto ad una variabile

Sia $f : \mathbb{R}^2 \to \mathbb{R}$, $z = f(x,y)$. Si dice *derivata parziale* della funzione f rispetto alla variabile x e si scrive

$$\frac{\partial f}{\partial x}$$

la derivata di f ottenuta considerando la funzione come dipendente solo dalla variabile x considerando le altre variabili come se fossero parametri. In modo analogo si definiscono le derivate parziali di una funzione rispetto alle altre variabili.

Esempio B.4 (Derivate parziali)

1. $f(x,y) = x^2 y$

$$\frac{\partial f}{\partial x} = 2xy \qquad \frac{\partial f}{\partial y} = x^2$$

2. $f(x,y) = \log(xy)$

$$\frac{\partial f}{\partial x} = \frac{1}{x} \qquad \frac{\partial f}{\partial y} = \frac{1}{y}$$

Definite le derivate parziali di una funzione rispetto ad una variabile, è ora possibile definire il cambio di variabile per le funzioni in due dimensioni.

Sia $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$, $(x,y) = (\Psi_1(x,y), \Psi_2(x,y))$; si dice *Jacobiano* della funzione di trasformazione Ψ e si indica con la notazione J_Ψ la matrice così definita:

$$J_\Psi = \begin{pmatrix} \dfrac{\partial \Psi_1}{\partial x} & \dfrac{\partial \Psi_1}{\partial y} \\ \\ \dfrac{\partial \Psi_2}{\partial x} & \dfrac{\partial \Psi_2}{\partial y} \end{pmatrix}$$

Un cambiamento di coordinate in $\mathbb{R}^2$ è dato da una funzione

$$\Psi : \mathbb{R}^2 \to \mathbb{R}^2$$

$$(u,v) \mapsto (x,y)$$

con determinate proprietà di regolarità (diffeomorfismo). Per cambiare le coordinate negli integrali, si usa la regola:

$$\int\int_A f(x,y)\,\mathrm{d}x\mathrm{d}y = \int\int_{\Psi^{-1}(A)} f(\Psi(u,v))\,|\det J_\Psi|\,\mathrm{d}x\mathrm{d}y$$

$$\begin{array}{ccc} \mathbb{R}^2_{(u,v)} & \xrightarrow{\Psi} & \mathbb{R}^2_{(x,y)} \\ & \overset{f\circ\Psi}{\searrow} & \downarrow f \\ & & \mathbb{R} \end{array}$$

Esempio B.5 In questo esempio si richiama il calcolo della costante di normalizzazione della distribuzione normale standard in due dimensioni per portare un esempio del cambio di coordinate nella risoluzione di un integrale.

$$\int\int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}x\mathrm{d}y$$

Per calcolare questo integrale è necessario pasare alle coordinate polari:

$$x = \rho\cos\theta$$

$$y = \rho\sin\theta$$

$$(\theta,\rho) \xrightarrow{\Psi} (x,y) = (\rho\,\cos\theta,\ \rho\,\sin\theta)$$

$$J_\Psi = \begin{pmatrix} \frac{\partial}{\partial\theta}\rho\,\cos\theta & \frac{\partial}{\partial\rho}\rho\,\cos\theta \\ \frac{\partial}{\partial\theta}\rho\,\sin\theta & \frac{\partial}{\partial\rho}\rho\,\sin\theta \end{pmatrix} = \begin{pmatrix} -\rho\,\sin\theta & \cos\theta \\ \rho\,\cos\theta & \sin\theta \end{pmatrix}$$

Calcolando il determinante Jacobiano si ottiene

$$\det J_\Psi = -\rho\,(\sin^2\theta + \cos^2\theta) = -\rho$$

ovvero

$$|\det J_\Psi| = \rho$$

Ne segue che

$$\begin{aligned} \int\int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}x\mathrm{d}y &= \int_0^{+\infty} \mathrm{d}\rho \int_0^{2\pi} \rho\,e^{-\frac{1}{2}\rho^2}\,\mathrm{d}\theta \\ &= \int_0^{+\infty} \rho\,e^{-\frac{1}{2}\rho^2}\,\mathrm{d}\rho \int_0^{2\pi} \mathrm{d}\theta \\ &= 2\pi\,\underbrace{\left[-e^{-\frac{1}{2}\rho^2}\right]_0^{+\infty}}_{1} \\ &= 2\pi \end{aligned}$$

e quindi

$$\left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x\right)^2 = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x \int_{-\infty}^{+\infty} e^{-\frac{Y^2}{2}}\,\mathrm{d}y = \int\int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}x\mathrm{d}y = 2\pi$$

In conclusione

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x = \sqrt{2\pi}$$

B.4 Sintesi dell'appendice

1. **Teorema del cambio dell'ordine di integrazione (*Fubini-Tonelli*)**

$$\int \mathrm{d}x \int f(x,y)\,\mathrm{d}y = \int \mathrm{d}y \int f(x,y)\,\mathrm{d}x = \int\int f(x,y)\,\mathrm{d}x\mathrm{d}y$$

2. **Cambio di variabili** Dato il cambio di variabili $x_i = f_i(y_1,\ldots,y_n) \qquad i = 1,\ldots,n$ co corrispondente Jacobiano

$$J_f = \det \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

se $A \subset \mathbb{R}^n$, vale che

$$\int \int_A \Psi(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n = \int \int_{f^{-1}(A)} \Psi(f(y)) \, |J_f| \, \mathrm{d}y$$

Inoltre, per determinare i nuovi estremi di integrazione, si utilizza il *metodo delle rette normali*, alcuni esempi del quale si possono osservare nelle figure riportate qui sotto.
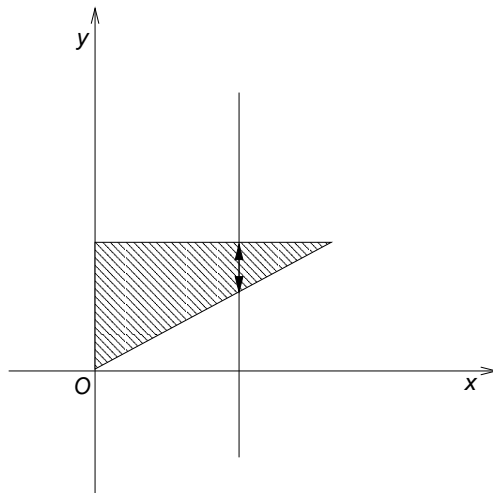
Figura B.5: Metodo delle rette normali quando varia la x.
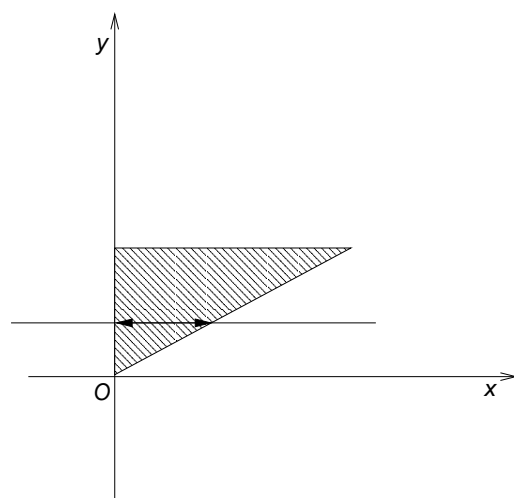
Figura B.6: Metodo delle rette normali quando varia la y.

Appendice C

Elementi di calcolo combinatorio

Si consideri un insieme $\Omega = \{a_1, \ldots, a_n\}$ contenente n elementi. Si ricorda che il simbolo $\binom{n}{r}$ si dice *coefficiente binomiale* e vale che $\binom{n}{r} = \dfrac{n!}{r!n-r!}$.

C.1 Disposizioni

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi con ripetizione e tenendo conto dell'ordine, ovvero il numero delle *disposizioni* di r elementi su n. Si hanno

$$1^o \text{ elemento} \longrightarrow n \text{ scelte}$$

$$2^o \text{ elemento} \longrightarrow n \text{ scelte}$$

$$\begin{matrix} . & & . \\ . & & . \\ . & & . \end{matrix}$$

$$r^o \text{ elemento} \longrightarrow n \text{ scelte}$$

In totale, le disposizioni sono $n \cdot n \cdots n = n^r$. Esse danno il numero di funzioni da un insieme di r elementi in un insieme di n elementi.

C.2 Disposizioni semplici

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi senza ripetizione e tenendo conto dell'ordine, ovvero il numero delle *disposizioni semplici* di r elementi su n. Si hanno

$$1^o \text{ elemento} \longrightarrow n \text{ scelte}$$

$$2^o \text{ elemento} \longrightarrow (n-1) \text{ scelte}$$

$$3^o \text{ elemento} \longrightarrow (n-2) \text{ scelte}$$

$$\vdots \qquad \vdots$$

$$r^o \text{ elemento} \longrightarrow (n-r+1) \text{ scelte}$$

In totale, le disposizioni semplici sono $n \cdot (n-1) \cdots (n-r+1) = \dfrac{n!}{n-r!}$ e si indicano con il simbolo D_r^n oppure $(n)_r$. Esse danno il numero di funzioni iniettive da un insieme di r elementi in un insieme di n elementi.
Se $r = n$, si parla di *permutazioni*.

C.3 Combinazioni semplici

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi senza ripetizione e senza tener conto dell'ordine, ovvero il numero delle *combinazioni semplici* di r elementi su n. Data una combinazione semplice di r elementi su n, si ottengono $r!$ disposizioni permutando gli r elementi. Il numero delle combinazioni è allora

$$\frac{1}{r!} D_r^n = \frac{n!}{r! n-r!} = \binom{n}{r}$$

Esse danno il numero di funzioni iniettive da un insieme di r elementi in un insieme di n elementi con immagine diversa.

C.4 Combinazioni

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi con ripetizione e senza tener conto dell'ordine, ovvero il numero delle *combinazioni* di r elementi su n. Data una combinazione $\{a_1, \ldots, a_n\}$, senza perdere di generalità si può supporre $a_1 \leq \cdots \leq a_n$. Si costruisce a partire da essa una combinazione semplice di r elementi in $n+r-1$ elementi nel modo seguente:

$$
\begin{aligned}
b_1 &= a_1 \\
b_2 &= a_2 + 1 \\
&\vdots \\
b_r &= a_r + r - 1
\end{aligned}
$$

Viceversa, data una combinazione semplice si può ad essa associare una combinazione. Le r-combinazioni sono quindi tante quante le r-combinazioni semplici in $n+r-1$, ovvero sono $\binom{n+r-1}{r}$.

C.5 Coefficiente multinomiale

Il numero di modi di formare k gruppi di $r_1, \ldots, r_k$ elementi ciascuno in modo tale che $r_1 + \cdots + r_k = n$ è dato dal *coefficiente multinomiale*

$$
\frac{n!}{r_1! r_2! \cdots r_k!}
$$

Per formare il primo gruppo di r_1 elementi, vi sono $\binom{n}{r_1}$ modi. Per il secondo gruppo, esso si può formare scegliendo gli elementi in $\binom{n-r_1}{r_2}$ modi. Si procede analogamente per formare i gruppi restanti. Si ottiene

$$
\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-\cdots-r_{k-1}}{r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}
$$

Appendice D

La formula di Stirling

La formula di Stirling ci dà il comportamento asintotico di $n!$ al crescere di n. Vale infatti che

$$\textbf{formula di Stirling:} \quad n! = \sqrt{2\pi}\; n^{n+\frac{1}{2}}\; e^{-n}(1 + 0(n^{-1}))$$

Per dimostrare tale formula, si prova in questa appendice un risultato più generale. Consideriamo la funzione Gamma di Eulero data da

$$\textbf{funzione Gamma:} \quad \Gamma(\alpha) = \int_0^{+\infty} x^\alpha e^{-x} dx$$

dove $\alpha \in \mathbb{R}$. Essa rappresenta una generalizzazione del fattoriale in quanto per ogni $\alpha \in \mathbb{R}$ vale che

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$

come si verifica facilmente integrando per parti. Se α è un numero naturale, per iterazione si ottiene

$$\Gamma(n + 1) = n!.$$

Per dimostrare la formula di Stirling si prova quindi il risultato più generale

$$\Gamma(\alpha + 1) = \sqrt{2\pi}\; \alpha^{\alpha+\frac{1}{2}}\; e^{-\alpha}(1 + 0(\alpha^{-1}))$$

Consideriamo il logaritmo della funzione integranda $\phi(x) = \log x^\alpha e^{-x} = \alpha \log x - x$. Facciamo lo sviluppo di Taylor di $\phi(x)$ nel punto di massimo α:

$$\phi(x) = \alpha \log \alpha - \alpha - \frac{1}{2\alpha}(x - \alpha)^2 + \sum_{k=3}^{n} \frac{(-1)^{k-1}}{k} \frac{(x - \alpha)^k}{\alpha^{k-2}} + \alpha \frac{(-1)^n}{n + 1} \frac{(x - \alpha)^{n+1}}{\xi^{n+1}}$$

dove $\xi \in [\alpha, x]$. Effettuiamo nell'integrale il cambio di variabile

$$u = \frac{x - \alpha}{\sqrt{\alpha}}, \qquad dx = \sqrt{\alpha}\; du$$

Si ottiene

$$\Gamma(\alpha+1)=\alpha^{\alpha+\frac{1}{2}}\ e^{-\alpha}\int_{-\sqrt{\alpha}}^{+\infty}e^{-\frac{u^2}{2}+\psi(u)}du$$

dove $\psi(u)=\sum_{k=3}^{n}\frac{(-1)^{k-1}}{k}\frac{u^k}{\alpha^{\frac{n}{2}-1}}+\alpha^{\frac{n+3}{2}}\frac{(-1)^n}{n+1}\frac{u^{n+1}}{(\alpha+\xi)^{n+1}}$ con $\xi\in[\alpha,u]$.
Dividiamo l'integrale in tre parti:

$$I_1=[-\sqrt{\alpha},-\alpha^\delta],\quad I_2=[-\alpha^\delta,\alpha^\delta],\quad I_3=[\alpha^\delta,+\infty]$$

dove $\delta>0$ è una costante opportunamente piccola. Per quanto riguarda I_1, I_2 osserviamo che la funzione $\phi(u)$ è concava e quindi anche la funzione $\theta(u)=-\frac{u^2}{2}+\psi(u)$ che si ottiene da ϕ con l'aggiunta di una costante ed un cambiamento lineare di variabile. Per $u\leq-\alpha^\delta$ abbiamo quindi $\theta(u)\leq-\frac{u}{\alpha^\delta}\theta(-\alpha^\delta)$ e per $u\geq-\alpha^\delta$ $\theta(u)\leq\frac{u}{\alpha^\delta}\theta(\alpha^\delta)$.
Dall'espansione di $\psi(u)$ con $n=2$ vediamo che per α abbastanza grande e $\delta<\frac{1}{6}$ abbiamo $\theta(-\alpha^\delta)<-\frac{\alpha^{2\delta}}{4}$, $\theta(\alpha^\delta)<-\frac{\alpha^{2\delta}}{4}$ e quindi vale che per $|u|\geq\alpha^{2\delta}$ $\theta(u)\leq-|u|\ \frac{\alpha^{2\delta}}{4}$. Ne segue che

$$\int_{I_1}e^{\theta(u)}du+\int_{I_3}e^{\theta(u)}du\leq\int_{|u|\geq\alpha^\delta}e^{-|u|\frac{\alpha^\delta}{4}}du=$$
$$\left[-\frac{8}{\alpha^\delta}e^{-|u|\frac{\alpha^\delta}{4}}\right]_{\alpha^\delta}^{+\infty}=-\frac{8}{\alpha^\delta}e^{-\frac{\alpha^{2\delta}}{4}}$$

Consideriamo ora I_2. Se scegliamo $n=3$ si ottiene:

$$e^{\psi(u)}=\exp\frac{1}{2}\frac{u^3}{\alpha^{\frac{1}{2}}}-\frac{\alpha^3}{\xi^4}u^4=1+\frac{1}{3}\frac{u^3}{\alpha^{\frac{1}{2}}}+0(\frac{u^4}{\alpha})$$

con $\xi\in[0,u]\subset I_2$ e per $|u|<\alpha^\delta$. Ne segue che

$$\int_{I_2}e^{-\frac{u^2}{2}+\psi(u)}du=\int e^{-\frac{u^2}{2}}du-\int_{I_2^c}e^{-\frac{u^2}{2}}du+\frac{\alpha^{-\frac{1}{2}}}{3}+\int_{I_2}u^3e^{-\frac{u^2}{2}}du+0(\alpha^{-1})=\sqrt{2\pi}+0(\alpha^{-1})$$

e quindi

$$\Gamma(\alpha+1)=\sqrt{2\pi}\ \alpha^{\alpha+\frac{1}{2}}\ e^{-\alpha}(1+0(\alpha^{-1}))$$

Appendice E

Dalle distribuzioni discrete a quelle assolutamente continue

Di sotto viene riportata una tabella che mette in evidenza le analogie che si possono individuare confrontando la teoria delle distribuzioni discrete con quella delle distribuzioni assolutamente contine.

C. Discreto		**C. Ass. Continuo**
Probabilità		Densità
$P(X = x)$	$\longrightarrow$	$f(x)$
Funzione di	ripartizione	$P(X \leq x)$
$\displaystyle\sum_{i \in I(X), i \leq x} P(X = i)$	$\longrightarrow$	$\displaystyle\int_{-\infty}^{x} f(s)\,\mathrm{d}s$
	Previsione di X	
$\displaystyle\sum_{i \in I(X)} i\,P(X = i)$	$\longrightarrow$	$\displaystyle\int_{-\infty}^{+\infty} s\,f(s)\,\mathrm{d}s$
Previsione di	$Y = \Psi(X)$	
$\displaystyle\sum_{i \in I(X)} \Psi(i)\,P(X = i)$	$\longrightarrow$	$\displaystyle\int_{-\infty}^{+\infty} \Psi(s)\,f(s)\,\mathrm{d}s$
	$P(X \in A)$	
$\displaystyle\sum_{i \in I(X), i \in A} P(X = i)$	$\longrightarrow$	$\displaystyle\int_{A} f(s)\,\mathrm{d}s$

Appendice F

Schema delle principali distribuzioni di probabilità

F.1 Distribuzioni discrete

Distribuzione	$I(X)$	$\mathbb{P}(X=k)$	$\mathbb{P}(X)$	$\sigma^2(X)$
Bernoulli p	$\{0,1\}$	$P(X=1)=p$	p	$p(1-p)$
Binomiale $B(n,p)$	$\{0,\dots,n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np\,(1-p)$
Geometrica p	$\{1,2,\dots\}$	$p\,(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Ipergeometrica (n,N,b)	$\{n-n\wedge(N-b),\dots,n\vee b\}$	$\frac{\binom{b}{k}\binom{N-b}{n-k}}{\binom{N}{n}}$	$n\frac{b}{N}$	$n\left(\frac{n-1}{N-1}\right)\frac{b}{N}\left(1-\frac{b}{N}\right)$
Poisson λ	$\mathcal{N}$	$\frac{\lambda}{k!}e^{-\lambda}$	λ	λ

F.2 Distribuzioni assolutamente continue

Distribuzione	$I(X)$	Densità	$P(X)$	$\sigma^2(X)$
Uniforme $[a,b]$	$[a,b]^a$	$\frac{1}{b-a}\, I_{[a,b]}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Esponenziale λ	$\mathbb{R}^+$	$\lambda\, e^{-\lambda x}\, I_{\{x\geq 0\}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale Std. $N(0,1)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi}}\, e^{\frac{x^2}{2}}$	0	1
Normale Gen. $N(\mu,\sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma $\Gamma(\alpha,\beta)$	$\mathbb{R}^+$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, I_{\{x\geq 0\}}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Beta $\beta(\alpha,\beta)$	$[0,1]$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$

[a] $b > a$

Appendice G

La distribuzione normale n-dimensionale

Densità	$f(x_1, \dots, x_n) = k\, e^{(-\frac{1}{2}Ax\cdot x + b\cdot x)}$ $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, A \in \mathcal{S}(n), b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$
Costante di Normalizzazione	$k = \sqrt{\frac{\det A}{(2\pi)^n}}\, e^{-\frac{1}{2}A^{-1}b\cdot b}$
Previsione	$P(X) = A^{-1}b \Rightarrow P(X_i) = (A^{-1}b)_i$
Matrice di Varianza e Covarianza	$C = A^{-1}$
Distribuzione marginale di X_i $(i = 1, \cdots, n)$	$X_i \sim N\left((A^{-1}b)_i, [A^{-1}]_{ii}\right)$

Appendice H

Il teorema di De Moivre-Laplace

Teorema H.1 *Sia* $(X_i)_{i \in \mathcal{N}}$ *una successione di numeri aleatori con distribuzione binomiale di parametri rispettivamente* $\mathcal{B}(n,p)$*. Dati i numeri aleatori* standardizzati

$$X_n^* = \frac{X_n - \mathbb{P}(X_n)}{\sigma(X_n)} = \frac{X_n - np}{\sqrt{np\tilde{p}}}$$

dove si è posto $\tilde{p} = 1 - p$*, vale che*

$$\mathbb{P}(|X_n^*| = x) = \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)}$$

dove $h_n = \dfrac{1}{\sqrt{np\tilde{p}}}$ *ed* $E_n(x)$ *è l'errore che tende a zero uniformemente se* x *è limitato.*

Dimostrazione. Se l'insieme dei valori possibili di X_n è $I(X_n) = \{0, 1, \ldots, n\}$, si ottiene che

$$I(X_n^*) = \{h_n(-np), h_n(1-np), \ldots, h_n(n-np)\}$$

dove $h_n = \dfrac{1}{\sqrt{np\tilde{p}}}$ è la spaziatura dei valori di X_n^*.
Definiamo $\phi_n(x) = \log \mathbb{P}(X_n^* = x)$ e consideriamo il suo rapporto incrementale:

$$\frac{\phi_n(x + h_n) - \phi_n(x)}{h_n} = \frac{1}{h_n} \log \frac{\mathbb{P}(X_n^* = x + h_n)}{\mathbb{P}(X_n^* = x)}$$

Posto $k = np + x\sqrt{np\tilde{p}}$, si ottiene

$$\begin{aligned}
\frac{1}{h_n}\log\frac{\mathbb{P}(X_n^*=x+h_n)}{\mathbb{P}(X_n^*=x)} &= \\
\frac{1}{h_n}\log\frac{(n-k)}{k+1}\frac{p}{\tilde{p}} &= \\
\sqrt{np\tilde{p}}\log\frac{n\tilde{p}-x\sqrt{np\tilde{p}}}{np+1+x\sqrt{np\tilde{p}}}\frac{p}{\tilde{p}} &= \\
\sqrt{np\tilde{p}}\log\frac{1-x\sqrt{\dfrac{p}{n\tilde{p}}}}{1+\dfrac{1}{np}+x\sqrt{\dfrac{\tilde{p}}{np}}}
\end{aligned}$$

Se n tende all'infinito, si può usare l'approssimazione $\log(1+x)=x+0(x)$ che vale per x sufficientemente piccolo. Si ottiene

$$\begin{aligned}
\sqrt{np\tilde{p}}\log\frac{1-x\sqrt{\dfrac{p}{n\tilde{p}}}}{1+\dfrac{1}{np}+x\sqrt{\dfrac{\tilde{p}}{np}}} &= \\
\sqrt{np\tilde{p}}\left[-x\sqrt{\frac{p}{n\tilde{p}}}+o(\frac{x^2}{n})-x\sqrt{\frac{p}{n\tilde{p}}}+o(\frac{x^2+1}{n})\right] &= \\
-xp-x\tilde{p}+o(\frac{x^2+1}{\sqrt{n}}) &= \\
-x+o(\frac{x^2+1}{\sqrt{n}})
\end{aligned}$$

La funzione $\phi_n(x)$ non è definita ovunque, ma solo per i valori di x tali che $\mathbb{P}(X_n^*=x)\neq 0$. Si estende ad ogni $x\in\mathbb{R}$ usando la rappresentazione

$$\phi_n(x)=\phi_n(0)+\int_0^x\phi_n'(y)dy$$

Se $x\leq y\leq x+h_n$, $\phi_n'(y)=\Delta_{h_n}\phi_n(x)=-x+o(\frac{x^2+1}{\sqrt{n}})=-y+o(\frac{x^2+1}{\sqrt{n}})$ da cui:

$$\begin{aligned}
\phi_n(x)=\phi_n(0)+\int_0^x\phi_n'(y)dy &= \\
\phi_n(0)+\int_0^x(-y)dy+o(\frac{x^3+x}{\sqrt{n}}) &= \\
\phi_n(0)-\frac{x^2}{2}+o(\frac{x^3+x}{\sqrt{n}})
\end{aligned}$$

Poiché $\phi_n(x) = \log \mathbb{P}(X_n^* = x)$, si ottiene

$$\log \mathbb{P}(X_n^* = x) = e^{\phi_n(0)} e^{-\frac{x^2}{2}} e^{E_n(x)}$$

dove $E_n(x) = o(\frac{x^3 + x}{\sqrt{n}})$.

Stimiamo $e^{\phi_n(0)}$ nel seguente modo. X_n^* è un numero aleatorio standardizzato, quindi con $\mathbb{P}(X_n^*) = 0$ e $\sigma^2(X_n^*) = 1$. Dalla disuguaglianza di Chebichev, si ha che:

$$\mathbb{P}(|X_n^*| \geq K) \leq \frac{1}{K^2}$$

Si può scegliere K in modo da rendere questa probabilità arbitrariamente piccola, ovvero per ogni $\epsilon > 0$ esiste K tale che

$$1 - \frac{1}{K^2} \leq \mathbb{P}(|X_n^*| < K) \leq 1$$

o anche

$$1 - \epsilon \leq \mathbb{P}(|X_n^*| < K) \leq 1$$

Poiché $\mathbb{P}(|X_n^*| < K) = \sum_{x,|x|<K} \mathbb{P}(|X_n^*| = x)$, ne segue

$$1 - \epsilon \leq \sum_{x,|x|<K} \mathbb{P}(|X_n^*| = x) \leq 1$$

$\mathbb{P}(|X_n^*| < K) = \sum_{x,|x|<K} \mathbb{P}(|X_n^*| = x) = \sum_{x,|x|<K} h_n e^{-\frac{x^2}{2}}$ è la somma di Riemann della funzione $e^{-\frac{x^2}{2}}$, quindi tende a $\int_{-K}^{K} e^{-\frac{x^2}{2}} dx$ per $n \to \infty$. Ne segue che

$$1 - \epsilon \leq \frac{e^{\phi_n(0)}}{h_n} \int_{-K}^{K} e^{-\frac{x^2}{2}} dx \leq 1$$

Facendo tendere K all'infinito, si ottiene

$$1 - \epsilon \leq \frac{e^{\phi_n(0)}}{h_n} \sqrt{2\pi} \leq 1$$

ovvero

$$\frac{e^{\phi_n(0)}}{h_n} \sqrt{2\pi} \longrightarrow_{n \longrightarrow \infty} 1$$

Ne segue che

$$\mathbb{P}(|X_n^*| = x) = \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)}$$

dove $E_n(x)$ è l'errore che tende a zero uniformemente se x è limitato. $\square$

Come applicazione del teorema, si ottiene un'approssimazione della funzione di ripartizione della distribuzione binomiale. Dati $a, b, 0 < a < b$, si calcola:

$$\mathbb{P}(a \leq X_n^* \leq b) = \sum_{a \leq x \leq b} \mathbb{P}(|X_n^*| = x) = \sum_{a \leq x \leq b} \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)}$$

A meno del resto, $\displaystyle\sum_{a \leq x \leq b} \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)}$ è la somma di Riemann di $e^{-\frac{x^2}{2}}$, quindi converge a $\displaystyle\int_a^b e^{-\frac{x^2}{2}} dx = \mathcal{N}(b) - \mathcal{N}(a)$, dove $\mathcal{N}(x)$ è la funzione di ripartizione della gaussiana standard. Si può quindi studiare la convergenza della funzione di ripartizione $F_n(x)$ di X_n^* nel seguente modo:

$$F_n(x) = \mathbb{P}(X_n^* \leq x) = \mathbb{P}(-k < X_n^* \leq x) + \mathbb{P}(X_n^* \leq -k) = \mathcal{N}(x) - \mathcal{N}(-k) + \mathbb{P}(X_n^* \leq -k)$$

Per la disuguaglianza di Chebichev, si può rendere $\mathbb{P}(X_n^* \leq -k)$ piccola a piacere. Inoltre anche $\mathcal{N}(-k)$ tende a zero per $k \longrightarrow \infty$, quindi la funzione di ripartizione della binomiale tende alla funzione di ripartizione della gaussiana standard.